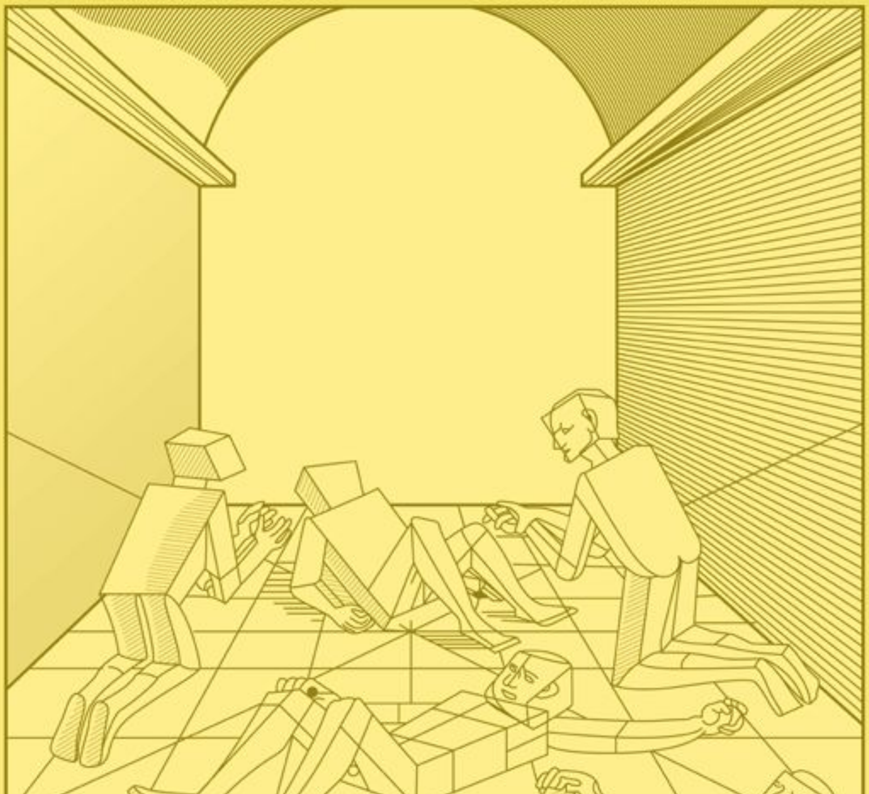
János Pach *Editor*

Thirty Essays on Geometric Graph Theory



Springer

Thirty Essays on Geometric Graph Theory

János Pach
Editor

Thirty Essays on Geometric Graph Theory

Springer

Editor
János Pach
Alfréd Rényi Institute of Mathematics
Hungarian Academy of Sciences
Budapest, Hungary

and

Chair of Combinatorial Geometry
École Polytechnique Fédérale de Lausanne
Lausanne, Switzerland

ISBN 978-1-4614-0109-4 ISBN 978-1-4614-0110-0 (eBook)
DOI 10.1007/978-1-4614-0110-0
Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012952159

Mathematics Subject Classification: 05-06, 05C62, 68R10, 05C35, 52C10

Cover image: Redrawn image of Erhard Schön's "Fünf Figuren in einem Gebäude," published in Unnderweissung der Proportzion (Nürnberg, 1538).

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Contributors

Bernardo M. Ábrego Department of Mathematics, California State University at Northridge, Los Angeles, CA, USA

Eyal Ackerman Department of Mathematics, Physics, and Computer Science, University of Haifa at Oranim, Tivon, Israel

Marwan Al-Jubeh Department of Computer Science, Tufts University, Medford, MA, USA

Greg Aloupis Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium

Brad Ballinger Department of Mathematics, Humboldt State University, Arcata, CA, USA

Luis F. Barba Instituto de Matemáticas, Universidad Nacional Autónoma de México, D.F., Mexico

Gill Barequet Department of Computer Science, Technion, Haifa, Israel

Department of Computer Science, Tufts University, Medford, MA, USA

Giuseppe Di Battista Dipartimento di Informatica e Automazione, Roma Tre University, Rome, Italy

Sarit Buzaglo Mathematics Department, Technion-Israel Institute of Technology, Haifa, Israel

Javier Cano Instituto de Matemáticas, Universidad Nacional Autónoma de México, D.F., Mexico

Josef Cibulka Department of Applied Mathematics, Charles University, Prague, Czech Republic

Sébastien Collette Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium

Walter Didimo Dipartimento di Ingegneria Elettronica e dell'Informazione, Università degli Studi di Perugia, Perugia, Italy

Adrian Dumitrescu Department of Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI, USA

Stefan Felsner Institut für Mathematik, Technische Universität Berlin, Berlin, Germany

Silvia Fernández-Merchant Department of Mathematics, California State University at Northridge, Los Angeles, CA, USA

Fabrizio Frati Chair of Combinatorial Geometry - École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Radoslav Fulek Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Michael Hoffmann Institute of Theoretical Computer Science, ETH Zürich, Zürich, Switzerland

Ferran Hurtado Departament de Matemàtica Aplicada II, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain

Mashhood Ishaque Department of Computer Science, Tufts University, Medford, MA, USA

Mikio Kano Department of Computer and Information Sciences, Ibaraki University, Hitachi, Ibaraki, Japan

Gyula Károlyi Institute of Mathematics, Eötvös University, Budapest, Hungary

Chaya Keller Einstein Institute of Mathematics, Hebrew University, Jerusalem, Israel

Alexandr V. Kostochka Department of Mathematics, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA

Jan Kynčl Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Prague, Czech Republic

Stefan Langerman Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium

Giuseppe Liotta Dipartimento di Ingegneria Elettronica e dell'Informazione Università degli Studi di Perugia, Perugia, Italy

Viola Mészáros Institute of Mathematics, Technical University of Berlin, Berlin, Germany

Kevin G. Milans Department of Mathematics, University of South Carolina, Columbia, SC, USA

Filip Morić Ecole Polytechnique Fédérale de Lausanne,Lausanne, Switzerland

János Pach Ecole Polytechnique Fédérale de Lausanne, Station 8, Lausanne 1015, Switzerland

Attila Pór Department of Mathematics, Western Kentucky University, Bowling Green, KY, USA

Michael J. Pelsmajer Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL, USA

Micha A. Perles Einstein Institute of Mathematics, Hebrew University, Jerusalem, Israel

Rom Pinchasi Mathematics Department, Technion-Israel Institute of Technology, Haifa, Israel

David Pritchard Centre for Education in Math and Computing, University of Waterloo, Waterloo, ON, Canada

Andrei M. Raigorodskii Mechanics and Mathematics Faculty, Department of Mathematical Statistics, and Random Processes, Moscow State University, Moscow, Russia

Faculty of Innovations and High Technology, Department of Data Analysis, Moscow Institute of Physics and Technology, Moscow, Russia

Yandex Research Laboratories, Moscow, Russia

Eduardo Rivera-Campo Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, México D.F., Mexico

Günter Rote Institut für Informatik, Freie Universität Berlin, Berlin, Germany

Noushin Saeedi Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

Toshinori Sakai Research Institute of Educational Development, Tokai University, Kanagawa, Japan

Gelasio Salazar Instituto de Física, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico

Deniz Sariöz The Graduate School and University Center of The City University of New York, New York, NY, USA

Marcus Schaefer Department of Computer Science, DePaul University, Chicago, IL, USA

André Schulz Institut für Mathematische Logik und Grundlagenforschung, Universität Münster, Münster, Germany

Micha Sharir School of Computer Science, Tel Aviv University, Tel Aviv, Israel

Courant Institute of Mathematical Sciences, New York University, New York, NY, USA

Adam Sheffer School of Computer Science, Tel Aviv University, Tel Aviv, Israel

Clifford Smyth Department of Mathematics and Statistics, University of North Carolina Greensboro, Greensboro, NC, USA

Diane L. Souvaine Department of Computer Science, Tufts University, Medford, MA, USA

Daniel Štefankovič Computer Science Department, University of Rochester, Rochester, NY, USA

Rudolf Stolař Department of Applied Mathematics, Charles University, Prague, Czech Republic

Andrew Suk Department of Mathematics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA

Kazuhiro Suzuki Department of Information Science, Kochi University, Kochi, Japan

Konrad J. Swanepoel Department of Mathematics, London School of Economics and Political Science, London, UK

Martin Tancer Faculty of Mathematics and Physics, Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Prague, Czech Republic

Gábor Tardos Rényi Institute, Budapest, Hungary

School of Computing Science, Simon Fraser University, Burnaby, BC, Canada

Csaba D. Tóth Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada

Department of Computer Science, Tufts University, Medford, MA, USA

Géza Tóth Rényi Institute, Hungarian Academy of Sciences, Budapest, Hungary

Jorge Urrutia Instituto de Matemáticas, Universidad Nacional Autónoma de México, D.F., Mexico

Virginia Urrutia-Galicia Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa Av. San Rafael Atlixco 186, México D.F. 09340, Mexico

Pavel Valtr Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Prague, Czech Republic

Uli Wagner Institut für Theoretische Informatik, ETH Zürich, Zürich, Switzerland

Emo Welzl Institute of Theoretical Computer Science, ETH Zürich, Zürich, Switzerland

Andrew Winslow Department of Computer Science, Tufts University, Medford, MA, USA

David R. Wood Department of Mathematics and Statistics, University of Melbourne, Melbourne, Australia

Introduction

János Pach

In the mathematical literature, the term "geometric graph theory" is often used in a somewhat vague sense: to cover any area of graph theory in which geometric methods seem to be relevant to the study of graphs defined by geometric means. In the present volume, by a *geometric graph* we mean a graph drawn in the plane so that its vertices are represented by distinct points and its edges by (possibly crossing) straight-line segments between these points such that no edge passes through a vertex different from its endpoints. *Topological graphs* are defined analogously, except that their edges can be represented by simple Jordan arcs [17].

In this sense, the theory of geometric and topological graphs starts with the study of *planar graphs*, initiated by Euler around 1750. For a long time it appeared that planar graphs, that is, graphs that can be drawn without edge crossings, do not have many interesting properties; they offer little excitement for mathematicians. *Kuratowski* and *Wagner* found simple characterizations of planar graphs in terms of forbidden subdivisions and minors, and it follows from *Steinitz's* work on convex polytopes that every planar graph can be drawn by noncrossing straight-line edges, as a geometric graph. In other words, every planar graph can be "stretched." This fact is usually referred to as *Fáry's theorem* [9]. One of the first really surprising results on planar graphs was the *Hanani–Tutte theorem* [7, 21], which states that if a graph can be drawn in such a way that any pair of its edges cross an even number of times, then it can also be drawn without edge crossing; that is, it must be a planar graph! It turns out that the reason why parity plays a role here lies in the *Jordan curve theorem*: Any curve connecting two points, both of which belong to the interior of a closed Jordan curve, must cross this curve an even number of times.

In the past quarter of a century, partially driven by the needs of computer graphics and other techniques of visualization, *graph drawing* has become a separate new area of research on the borderline of graph theory and computational geometry,

J. Pach (✉)
Ecole Polytechnique Fédérale de Lausanne, Station 8, Lausanne 1015, Switzerland
e-mail: pach@renyi.hu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_1,

with its annual international symposia and regular conference proceedings. The first such conference took place in June 1992, in Marino (near Rome). The subject has developed in close cooperation with industrial researchers developing software for visualization. Many interesting mathematical questions were asked, which were clearly motivated by potential applications. For instance, what is the size of the smallest integer grid such that every planar graph of n vertices admits a crossing-free straight-line drawing in which every vertex is mapped to a grid point [10]? Does there exist a positive-valued function $a(d)$ such that every planar graph of maximum degree d admits a crossing-free straight-line drawing, in which the angle between any two adjacent edges is at least $a(d)$ [16]? Many other more realistic measures of resolution were also considered. Geometric and topological graph theory has become one of the theoretical pillars of graph drawing.

Most graphs G are not planar. Nevertheless, we often have to represent them in the plane, and we may want to minimize the number of crossings in the resulting drawing. The smallest number of crossings that we can achieve is called the *crossing number* of G. *Turán's* famous "brick factory problem" asks for the crossing number of a complete bipartite graph with n vertices in each of its vertex classes [20]. In spite of many attempts to solve this problem, we still do not have even an asymptotically tight answer to this question. According to *Zarankiewicz's conjecture* [12], this quantity is equal to $\left(\frac{1}{16}+o(1)\right)n^4$. Many interesting related problems can be raised. For example, one can define the *pair-crossing number* of G as the minimum number of crossing pairs of edges over all possible drawings of G [8, 19]. Do the crossing number and the pair-crossing number coincide for every graph? We know that if we restrict our attention to straight-line drawings of G, we obtain a new parameter, different from the crossing number. The minimum number of crossings over all straight-line drawings of G is called the *linear crossing number* of G. It is known that there are graphs with crossing number 4 and with arbitrarily large linear crossing numbers [6]. According to a particularly useful inequality of Leighton [15] and, independently, of Ajtai et al. [3], the crossing number of any graph with n vertices and $e > 3n-6$ edges is at least a positive constant times e^3/n^2. Apart from the value of the constant, this bound is tight. It has found many interesting applications in combinatorial geometry and number theory.

A topological graph is called a *thrackle* if any pair of its edges meet precisely once, either at an endpoint or at a proper crossing [22]. According to *Conway's* celebrated *thrackle conjecture*, every thrackle has at most as many edges as vertices. The conjecture is known to be true for straight-line thrackles (geometric graphs) and for thrackles that can be drawn in such a way that that every vertical line meets every edge in at most one point [13]. However, the best-known general upper bound for the number of edges of an n-vertex thrackle is only $1.42n$ (see [11]). The fact that this simply formulated puzzle has been open for almost half a century indicates how little we know about crossing patterns of edges in a topological graph. A topological graph is called *simple* if any pair of its edges meet at most once, either at an endpoint or at a proper crossing. Conway's conjecture can now be rephrased as follows: Every simple topological graph with n vertices and more than n edges has two disjoint edges (that is, two edges that do not cross and do not share an endpoint).

In the same spirit, *Erdős, Hanani, Kupitz, Perles,* and others raised a number of exciting extremal problems [4, 5, 14]. What is the maximum number of edges that a simple topological graph of n vertices can have if it contains no k pairwise disjoint edges for a fixed $k > 2$ or some other fixed forbidden configuration? There has been a lot of activity in this area, yet most of the above questions are still open. For instance, it is conjectured that, for any fixed k, the maximum number of edges of a (simple) topological graph with n vertices and no k pairwise crossing edges is $O(n)$. The conjecture holds for $k \leq 4$ (see [1, 2]). It would be true for every k if the answer to the following question of *Erdős* were in the affirmative: Does there exist a constant $\chi(k)$ for every $k \geq 3$ such that any system of segments in the plane, no k of which are pairwise crossing, can be colored by $\chi(k)$ colors so that no two segments that cross each other receive the same color? However, it has been proved that the answer is no even for $k = 3$.

The first conference dedicated to geometric graph theory was held at Rutgers University, New Jersey, in the framework of the DIMACS Special Focus on Discrete and Computational Geometry, in the Fall of 2002 (see [18]). The second such meeting took place eight years later, as part of the Special Semester on Discrete and Computational Geometry organized by the Bernoulli Center at EPFL, Lausanne. The progress in this area made during this period is striking. The present volume is a careful selection of 30 invited and thoroughly refereed survey and research articles reporting on significant recent achievements in geometric graph theory.

References

1. E. Ackerman, On the maximum number of edges in topological graphs with no four pairwise crossing edges. Discr. Comput. Geom. **41**, 365–375 (2009)
2. P. Agarwal, B. Aronov, J. Pach, R. Pollack, M. Sharir, Quasi-planar graphs have a linear number of edges. Combinatorica **17**, 1–9 (1997)
3. M. Ajtai, V. Chvátal, M. Newborn, E. Szemerédi, Crossing-free subgraphs. Ann. Discr. Math. **12**, 9–12 (1982)
4. N. Alon, P. Erdős, Disjoint edges in geometric graphs. Discr. Comput. Geom. **4**, 287–290 (1989)
5. S. Avital, H. Hanani, Graphs. Gilyonot Lematematika **3**, 2–8 (1966) [in Hebrew]
6. D. Bienstock, N. Dean, Bounds for rectilinear crossing numbers. J. Graph Theor. **17**, 333–348 (1993)
7. Ch. Chojnacki (A. Hanani), Über wesentlich unplättbare Kurven im dreidimensionalen Raume. Fund. Math. **23**, 135–142 (1934)
8. P. Erdős, R.K. Guy, Crossing number problems. Am. Math. Mon. **80**, 52–58 (1973)
9. I. Fáry, On straight line representation of planar graphs. Acta Univ. Szeged. Sect. Sci. Math. **11**, 229–233 (1948)
10. H. de Fraysseix, J. Pach, R. Pollack, How to draw a planar graph on a grid. Combinatorica **10**, 41–51 (1990)
11. R. Fulek, J. Pach, A computational approach to Conway's thrackle conjecture. Comput. Geom. **44**, 345–355 (2011)
12. R.K. Guy, The decline and fall of Zarankiewicz's theorem, in *Proof Techniques in Graph Theory* (Academic Press, New York, 1969), pp. 63–69

13. H. Hopf, E. Pannwitz, Aufgabe no. 167. Jahresbericht Deutschen Mathematiker-Vereinigung **43**, 114 (1934)
14. Y. Kupitz, Extremal problems in combinatorial geometry, in *Aarhus University Lecture Notes Series*, vol. 53 (Aarhus University, Aarhus, 1979)
15. T. Leighton, in *Complexity Issues in VLSI, Foundations of Computing Series* (MIT, Cambridge, 1983)
16. S. Malitz, A. Papakostas, On the angular resolution of planar graphs. SIAM J. Discr. Math. **7**, 172–183 (1994)
17. J. Pach, Geometric graph theory, in *Handbook of Discrete and Computational Geometry*, 2nd edn., ed. by J.E. Goodman, J. O'Rourke (Chapman & Hall/CRC, Boca Raton, FL, 2004), pp. 219–238 (Chap. 10)
18. J. Pach (ed.), *Towards a Theory of Geometric Graphs, Contemporary Mathematics*, vol. 342 (American Mathematical Society, Providence, RI, 2004)
19. J. Pach, G. Tóth, Which crossing number is it, anyway? J. Comb. Theor. Ser. B **80**, 225–246 (2000)
20. P. Turán, A note of welcome. J. Graph Theor. **1**, 7–9 (1977)
21. W.T. Tutte, Toward a theory of crossing numbers. J. Comb. Theor. **8**, 45–53 (1970)
22. D.R. Woodall, Thrackles and deadlock, in *Combinatorial Mathematics and Its Applications*, ed. by D.J.A. Welsh (Academic Press, London, 1969), pp. 335–348

The Rectilinear Crossing Number of K_n: Closing in (or Are We?)

Bernardo M. Ábrego, Silvia Fernández-Merchant, and Gelasio Salazar

Abstract The calculation of the rectilinear crossing number of complete graphs is an important open problem in combinatorial geometry, with important and fruitful connections to other classical problems. Our aim in this chapter is to survey the body of knowledge around this parameter.

Mathematics Subject Classification (2010): 52C30, 52C10, 52C45, 05C62, 68R10, 60D05, 52A22

1 Introduction

In a *rectilinear* (or *geometric*) drawing of a graph G, the vertices of G are represented by points, and an edge joining two vertices is represented by the straight segment joining the corresponding two points. Edges are allowed to cross, but an edge cannot contain a vertex other than its endpoints. The *rectilinear* (or *geometric*) crossing number $\overline{\mathrm{cr}}(G)$ of a graph G is the minimum number of pairwise crossinsgs of edges in a rectilinear drawing of G in the plane.

B.M. Ábrego (✉) • S. Fernández-Merchant
Department of Mathematics, California State University at Northridge, Northridge, CA, USA
e-mail: bernardo.abrego@csun.edu; silvia.fernandez@csun.edu

G. Salazar
Instituto de Física, Universidad Autónoma de San Luis Potosí, San Luis Potosí, Mexico
e-mail: gsalazar@ifisica.uaslp.mx

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_2,

1.1 The Relevance of $\overline{\text{cr}}(K_n)$

As with every graph theory parameter, there is a natural interest in calculating the rectilinear crossing number of certain families of graphs, such as the complete bipartite graphs $K_{m,n}$ and the complete graphs K_n. The estimation of $\overline{\text{cr}}(K_n)$ is of particular interest, since $\overline{\text{cr}}(K_n)$ equals the minimum number $\square(n)$ of convex quadrilaterals defined by n points in the plane in general position; the problem of determining $\square(n)$ belongs to a collection of classical combinatorial geometry problems, the so-called Erdős–Szekeres problems. For a comprehensive survey on results and open questions on these and related problems, we refer the reader to the monograph by Brass et al. [16].

Another important motivation to study $\overline{\text{cr}}(K_n)$ is its close connection with the celebrated Sylvester four-point problem from geometric probability. Sylvester asked what the probability is that four points chosen at random in the plane form a convex quadrilateral [29]. After it became clear that this is an ill-posed question [30], Sylvester put forward a related conjecture. Let R be a bounded convex open set in the plane with finite area, and let $q(R)$ be the probability that four points chosen randomly from R define a convex quadrilateral. Then (Sylvester's conjecture [20]) $q(R)$ is minimized when R is a circle or an ellipse, and maximized when R is a triangle. This conjecture was proved by Blashke in 1917 [15]. Scheinerman and Wilf addressed in [27] the general problem when R is not required to be convex. It is easy to see that in this case $q(R)$ can be made arbitrarily close to 1 by choosing R to be a very thin annulus. The remaining problem is to determine the infimum $q_* := \inf q(R)$, taken over all open sets R with finite area. Scheinerman and Wilf established the striking connection

$$q_* = \lim_{n\to\infty} \frac{\overline{\text{cr}}(K_n)}{\binom{n}{4}}, \tag{1}$$

thus inextricably linking the estimation of *Sylvester's four-point constant* q_* to the (asymptotic) behavior of $\overline{\text{cr}}(K_n)$.

As we shall see below, recent developments have unveiled a close relationship between $\overline{\text{cr}}(K_n)$ and yet another classical combinatorial geometry parameter: the number of $(\leq k)$-edges in an n-point set.

1.2 Purpose and Timeliness of This Survey

Up until 2000, very little was known about $\overline{\text{cr}}(K_n)$. Since then, our knowledge of this problem has seen a tremendous growth. Surprising and useful connections to other classical problems have been unveiled. The current estimates for $\overline{\text{cr}}(K_n)$ have reached a point that would have seemed unlikely (to say the least) at the beginning of the previous decade.

For instance, before 2000 the ratio between the best lower and upper bounds for q_* was about 0.755; at the time of writing this survey, this ratio has been raised above 0.998. The implied success in our understanding of the problem cannot be understated—hence, the words “closing in” in the title of this chapter. Moreover, as we have already mentioned above and shall see below in more detail, the problem of estimating $\overline{\mathrm{cr}}(K_n)$ has turned out to be intimately related to other classical combinatorial geometry problems. Nowadays, anyone seriously interested in $(\leq k)$-edges or in halving lines has no alternative but to take a careful look at the literature on $\overline{\mathrm{cr}}(K_n)$ that has been produced in the last seven or eight years.

On the more cautious side, we must also note that the steady progress achieved on the estimation of $\overline{\mathrm{cr}}(K_n)$, both from the lower and the upper bounds' fronts, seems to have reached an impasse. To a researcher not too familiar with the field, the ratio 0.998 mentioned in the previous paragraph might signal an imminent closure on the determination of q_*. This is by no means the prevalent feeling among most (if not all) researchers actively working on this problem. Hardly any relevant new insights have been reported for some time. This humbling reality prompted us to include a word of caution (“or are we?”) in the title of this chapter.

With this in mind, it makes sense to sit down and reflect on what has been done, to highlight the key developments, and to record the state of the art of the problem. We also see this as an opportunity to candidly (and, at times, informally) explain the obstacles that seem to prevent any further substantial progress with the current techniques, in the hopes that this will foster the development of refined or substantially novel techniques to attack this fundamental problem.

1.3 Structure of This Survey

The problem of estimating $\overline{\mathrm{cr}}(K_n)$ breaks into the two problems of establishing upper and lower bounds for this parameter, with the problem of finding exact values of $\overline{\mathrm{cr}}(K_n)$ lying, evidently, within both realms.

Before moving on to separate discussions on the problems of lower and upper bounding $\overline{\mathrm{cr}}(K_n)$, we shall review one of the main foundations behind our current knowledge of $\overline{\mathrm{cr}}(K_n)$. The Rectilinear Crossing Number (RCN) project, led by Aichholzer, has been a fruitful source of inspiration as well as an invaluable tool for establishing results and testing conjectures. In Sect. 2 we describe the nature and reach of the RCN project, which, as we will see, has both a claim and an impact on both the lower- and upper-bounding fronts.

In Sect. 3 we give an overview of the state of the art of the problem of lower bounding $\overline{\mathrm{cr}}(K_n)$ circa 2003.

Besides Aichholzer's RCN project, there seems to be a general consensus on the other main foundation behind our current knowledge of $\overline{\mathrm{cr}}(K_n)$. A major breakthrough was achieved around 2003, when two independent teams of researchers elucidated the close connection between $\overline{\mathrm{cr}}(K_n)$ and the number of $(\leq k)$-edges in an n-point set [4, 25]. A good estimate on the number of such $(\leq k)$-edges, also given in these papers, yielded an impressively improved lower bound on $\overline{\mathrm{cr}}(K_n)$. We devote Sect. 4 to a review of these cornerstone results.

In Sect. 5 we overview the subsequent efforts to refine the bounds for the number of $(\leq k)$-edges given in [4, 25], in the quest for improved lower bounds for $\overline{\text{cr}}(K_n)$.

In Sect. 6 we discuss the different approaches to establishing upper bounds for $\overline{\text{cr}}(K_n)$.

Section 7 contains a brief summary of the state of the art of the problem at the time of writing this survey. We present the current best estimates (lower and upper bounds) for q_*, as well as an annotated table with the values of n for which the exact value of $\overline{\text{cr}}(K_n)$ is known.

We conclude this survey by reflecting on some possible future developments around this fundamental problem. We discuss the difficulties that lie behind our current impasse, and outline a somewhat promising approach that may pave the way toward future improvements.

2 The RCN Project

Around 2000, a team of researchers led by Aichholzer undertook the task of building databases with all the distinct (up to order type equivalence; see below) n-point configurations in general position, for $n \leq 10$ [8, 12, 24]. The raw knowledge of all possible n-point configurations put Aichholzer and his collaborators in a position to explore in depth several classical combinatorial geometry problems. In particular, it allowed for the exact calculation of $\overline{\text{cr}}(K_n)$ for small values of n.

The criterion used by Aichholzer et al. to discriminate if two collections of points are nonisomorphic is based on the concept of order types. Consider an (ordered) n-point set $P = \{p_1, p_2, \ldots, p_n\}$ in general position. To each three integers i, j, k with $1 \leq i < j < k \leq n$, associate a *sign* (or *order type*) $\text{sign}(ijk)$ according to the following rule. If, as we traverse the triangle defined by p_i, p_j, and p_k by following the edges $\overline{p_i p_j}$, $\overline{p_j p_k}$, and $\overline{p_k p_i}$ in the given order, the resulting closed curve has a clockwise orientation, then let $\text{sign}(ijk) := +$. Otherwise, let $\text{sign}(ijk) := -$. The collection of the order types of all the triples of points of P is the *order type* of P. Now let Q be another n-point set in general position. If the elements of Q can be ordered $\{q_1, q_2, \ldots, q_n\}$ so that the order types of P and Q are the same, then P and Q are *order type equivalent* (under the mapping $p_i \mapsto q_i$ for $i = 1, 2, \ldots, n$). We simply say that P and Q *have the same order type*.

Order type equivalence is a natural isomorphism criterion for point sets in general position. For crossing number purposes, it is certainly the relevant paradigm. Indeed, suppose that P and Q have the same order type. Then there is a bijection from the points of P to the points of Q so that four points in P span a convex quadrilateral if and only if the corresponding four points in Q span a convex quadrilateral. Conversely, if this last condition holds, then P and Q have the same order type.

Aichholzer et al. constructed the complete database of all distinct order types on n points, for all $n \leq 10$. As an application, they verified that $\overline{\text{cr}}(K_{10}) = 62$ (this had been proved by Brodsky et al. in [17]).

Without building the complete database for $n = 11$, the information gathered by Aichholzer et al. for $n \leq 10$ allowed them to calculate $\overline{\text{cr}}(K_{11})$ and $\overline{\text{cr}}(K_{12})$.

To achieve this, taking their database for 10 points as a starting point, they analyzed (for $m = 10$, and then for $m = 11$) which m-point order types may possibly be extended to $(m+1)$-point sets that correspond to crossing-minimal drawings of K_{m+1}.

The determination of the rectilinear crossing numbers of K_{11} and K_{12} marks the beginning of the RCN project. As one of the major achievements of the RCN, Aichholzer developed some impressively accurate heuristics to generate geometric drawings of K_n with few crossings. Aichholzer set up a web page (http://www.ist.tugraz.at/staff/aichholzer/research/rp/triangulations/crossing/) to keep track of the best geometric drawings of K_n available, as well as of the number of distinct (up to order type equivalence) drawings achieving the current minimum.

The results reported by Aichholzer (http://www.ist.tugraz.at/staff/aichholzer/research/rp/triangulations/crossing/) have had a major lasting impact in the field. As new results and techniques to find improved lower bounds have become available (see Sects. 4 and 5), it has been possible to determine the exact value of $\overline{\text{cr}}(K_n)$ for more values of n (see Sect. 7). The outstanding quality of the upper bounds obtained by Aichholzer is evidenced by the fact that the drawings reported in Aichholzer (http://www.ist.tugraz.at/staff/aichholzer/research/rp/triangulations/crossing/) turned out to be crossing-optimal for all $n \leq 27$ and for $n = 30$ (for $n = 28$ and 29 the exact value of $\overline{\text{cr}}(K_n)$ is still unknown). At the time of writing this chapter, the best upper bounds available (see Sect. 6) are obtained from constructions that build upon "base" drawings of K_n for relatively small values of n. As further evidence of the influence of the RCN, we note that the base drawings used have been obtained by small modifications of drawings given in Aichholzer (http://www.ist.tugraz.at/staff/aichholzer/research/rp/triangulations/crossing/).

As a final note, we mention that Aichholzer and Krasser subsequently completed the database of all distinct order types of 11-point sets [13] (http://www.ist.tugraz.at/staff/aichholzer/research/rp/triangulations/ordertypes/). Using this database as a starting point, they were able to compute $\overline{\text{cr}}(K_n)$ for all $n \leq 17$. Building the complete database of all the order type nonequivalent 12-point sets seems to be an unfeasible task; not only it is estimated that the storage of these 12-point sets would require several petabytes of memory, but there are also some important technical difficulties.[1]

3 Lower Bounds I: Before 2004

In a paper published in 1972, Guy [22] gave the exact value of $\overline{\text{cr}}(K_n)$ for $n \leq 9$. Almost 30 years later, Brodsky et al. [17] pushed the existing techniques to their limit, and introduced some clever new arguments, to calculate the exact value of $\overline{\text{cr}}(K_{10})$.

[1] Aichholzer, personal communication.

As one of the first results of the RCN project (see Sect. 2), Aichholzer et al. [9] gave computer-assisted proofs that $\overline{\text{cr}}(K_{11}) = 102$ and $\overline{\text{cr}}(K_{12}) = 153$.

Because each of the n subsets of size $n-1$ of an n-point set P has at most $\overline{\text{cr}}(K_{n-1})$ crossings, and each crossing of P appears in exactly $n-4$ such subsets, it follows that $(n-4)\overline{\text{cr}}(K_n) \geq n\overline{\text{cr}}(K_{n-1})$. This is equivalent to

$$1 \geq \frac{\overline{\text{cr}}(K_n)}{\binom{n}{4}} \geq \frac{\overline{\text{cr}}(K_{n-1})}{\binom{n-1}{4}},$$

which shows that Sylvester's four-point constant q_* defined in (1) actually exists. Starting from a lower bound for $\overline{\text{cr}}(K_m)$ for any fixed m, one can obtain a lower bound for $\overline{\text{cr}}(K_n)$ for every $n > m$ (and consequently a lower bound for q_*) by iterating $\overline{\text{cr}}(K_n) \geq \lceil \overline{\text{cr}}(K_{n-1})n/(n-4) \rceil$. This technique was used by Brodsky et al. [17] with $\overline{\text{cr}}(K_{10}) = 62$ to show that $q_* > 0.3001$. Adding to this argument the fact that $\overline{\text{cr}}(K_n)$ and $\binom{n}{4}$ have the same parity when n is odd (this easily follows from (2) but was proved for any nonnecessarily rectilinear drawing of K_n by Eggleton and Guy [21]), and using $\overline{\text{cr}}(K_{11}) = 102$, Aichholzer et al. [9] showed that $q_* > 0.3115$.

Building upon ideas from Welzl [34] and Wagner and Welzl [32], Wagner [31] used a completely novel approach to show that $q_* > 0.3288$. Wagner's work is particularly significant, since it deviates from the traditional approach of lower bounding q_* by using a particular lower bound and a counting argument. Indeed, the ideas in [31] are prescient of the revolutionary approach that will be reviewed in the next section.

4 Lower Bounds II: The Breakthrough

Our understanding of geometric drawings of K_n underwent a phase transition by unveiling a close relationship with k-edges. We recall that if P is an n-point set, and $0 \leq k \leq n/2 - 1$, a *k-edge* of P is a line through two points of P leaving exactly k points on one side. A $(\leq k)$-edge is a j-edge with $j \leq k$. The number of k- and $(\leq k)$-edges of P are denoted by $E_k(P)$ and $E_{\leq k}(P)$, respectively. Finally, let $E_{\leq k}(n)$ denote the minimum $E_{\leq k}(P)$, taken over all n-point sets P in general position.

For an n-point set P in the plane in general position, let $\overline{\text{cr}}(P)$ denote the number of crossings in the rectilinear drawing of K_n induced by P. The following was proved independently by Lovász et al. [25], and by Ábrego and Fernández-Merchant [4]:

$$\overline{\text{cr}}(P) = \sum_{k=0}^{\lfloor n/2 \rfloor - 2} (n-2k-3)E_{\leq k}(P) - \frac{3}{4}\binom{n}{3} + (1+(-1)^{n+1})\frac{1}{8}\binom{n}{2}. \tag{2}$$

The relevance of this connection between $\overline{\text{cr}}(P)$ and $E_{\leq k}(P)$ was made evident in both [4, 25] by proving that

$$E_{\leq k}(n) \geq 3\binom{k+2}{2}, \quad \text{for } 0 \leq k \leq n/2 - 1. \tag{3}$$

Substituting (3) into (2) yields

$$\overline{\text{cr}}(K_n) \geq \frac{3}{8}\binom{n}{4} + \Theta(n^3), \tag{4}$$

thus implying the remarkably improved bound $q_* \geq 3/8 = 0.375$.

We recall that the *crossing number* $\text{cr}(G)$ of a graph G is the minimum number of pairwise crossings of edges in a (nonnecessarily geometric) drawing of G in the plane. There are drawings of K_n with exactly $\lambda_n := (1/4)\lfloor n/2 \rfloor \lfloor (n-1)/2 \rfloor \lfloor (n-2)/2 \rfloor \lfloor (n-3)/2 \rfloor$ crossings, and it is widely believed that these drawings are crossing-minimal; that is, it is conjectured that $\text{cr}(K_n) = \lambda_n$ for every positive integer n. This conjecture has been verified for $n \leq 12$ [22, 26]. Since $\text{cr}(K_n) \leq \lambda_n$, it follows at once that $\lim_{n\to\infty} \text{cr}(K_n)/\binom{n}{4} \leq 3/8$.

This last upper bound gives an additional significance to (4). With this motivation, Lovász et al. pushed a little further, invoking the following from [33]:

$$E_{\leq k}(n) \geq \binom{n}{2} - n\sqrt{n^2 - 2n - 4k^2 + 4k}. \tag{5}$$

This last bound is better than (3) for $k > 0.4956n$. Using (3) for $k \leq 0.4956n$, and (5) for $k > 0.4956n$, Lovász et al. derived the slightly improved bound $q_* > (3/8) + 10^{-5}$. Although numerically marginal, this improvement is significant because it shows that $\text{cr}(K_n)$ and $\overline{\text{cr}}(K_n)$ differ in the asymptotically relevant term.

5 Lower Bounds III: Further Improvements

Since the key connection (2) was proved in [4, 25], all subsequent efforts to lower bound $\overline{\text{cr}}(K_n)$ have been focused on finding better estimates for $E_{\leq k}(n)$.

The first improvement was reported in [14], giving a lower bound for $E_{\leq k}(n)$ that is strictly better than (3) for $k > 0.4651n$. The bound given in [14] is in terms of a complicated expression. For our current surveying purposes, it suffices to mention that using this bound, Balogh and Salazar proved that $\overline{\text{cr}}(K_n) > 0.37553\binom{n}{4} + \Theta(n^3)$.

Another significant improvement was achieved by Aichholzer et al. [10], who proved that

$$E_{\leq k}(n) \geq 3\binom{k+2}{2} + 3\binom{k+2-\lfloor n/3 \rfloor}{2} - \max\{0, (k+1-\lfloor n/3 \rfloor)(n - 3\lfloor n/3 \rfloor)\}. \tag{6}$$

A shorter proof of (6), given in the more general context of pseudolinear drawings, was given in [1].

Substituting (6) into (2), one obtains the improved estimate $q_* \geq 41/108 > 0.37962$. Moreover, it is possible to use the bound by Balogh and Salazar [14] in the range $k > 0.4864n$ to obtain the marginally better $q_* > 0.37968$.

The current best lower bound known for q_* is derived using a result recently reported by Ábrego et al. [3, 7]. They proved that for every k and n such that $\lceil (4n-11)/9 \rceil - 1 \le k \le (n-2)/2$,

$$E_{\le k}(n) \ge u_k(n) \ge \binom{n}{2} - \frac{1}{9}\sqrt{1-\frac{2k+2}{n}}(5n^2+19n-31). \tag{7}$$

The function u_k is asymptotic to the latter expression and it is better than all previous bounds [including (5) (6), and the bound in [14]] across its full range $\lceil (4n-11)/9 \rceil \le k \le (n-2)/2$. In addition, Ábrego et al. [3] constructed point sets achieving equality on (6) for all $k < \lceil (4n-11)/9 \rceil$. Using (6) for $k < \lceil (4n-11)/9 \rceil$, and (7) for $\lceil (4n-11)/9 \rceil \le k \le (n-2)/2$. It follows from (2) that $\overline{\text{cr}}(K_n) \ge (277/729)\binom{n}{4} + \Theta(n^3)$, thus implying that $q_* \ge 277/729 > 0.37997$.

6 Upper Bounds

The literature on crossing numbers of particular families of graphs is vastly dominated by papers that focus on establishing lower bounds. Most of the time, a natural drawing suggests itself with relatively little effort. When successive attempts to produce better drawings fail, this is seen as plausible evidence that the proposed drawing is indeed optimal. The efforts are then directed in the opposite, and usually remarkably harder, direction: proving nontrivial lower bounds for the crossing numbers of the graphs upon consideration.

The problem of upper bounding the rectilinear crossing number of K_n is a notable exception to this trend. The goal is to describe a way to draw K_n with as few crossings as possible, for arbitrarily large values of n, so as to have at least an educated guess at the asymptotic value $q_* = \lim_{n\to\infty} \overline{\text{cr}}(K_n)/\binom{n}{4}$. Over the years, several strategies to draw K_n with few crossings have been put forward. However, to this day there has not been a clear candidate for an optimal drawing. The only common characteristic is that almost all drawings with few crossings have (or are really close to have) threefold symmetry with respect to a point. That is, the underlying point set P of the drawing is partitioned into three sets (we call them *wings*) of size $n/3$ each, with the property that rotating each wing $2\pi/3$ and $4\pi/3$ around a suitable point generates the other two wings.

In the early 1970s, Jensen [23] was the first to propose a way to draw K_n for arbitrarily large values of n. His construction gave specific coordinates for $n/3$ points in a wing, and then he obtained the remaining two wings by rotating $2\pi/3$ and $4\pi/3$ around the origin. As a result, he obtained $q_* \le 7/18 < 0.38889$.

At around the same time, Singer [28] started the trend of recursively constructing drawings of K_n. His idea was to start with a good drawing of $K_{n/3}$, apply an affine transformation to it to make the slope of each of its edges sufficiently close to zero, and then add the $2\pi/3$ and $4\pi/3$ rotations of the resulting drawing to obtain the other two wings (see Fig. 1a). This construction shows that

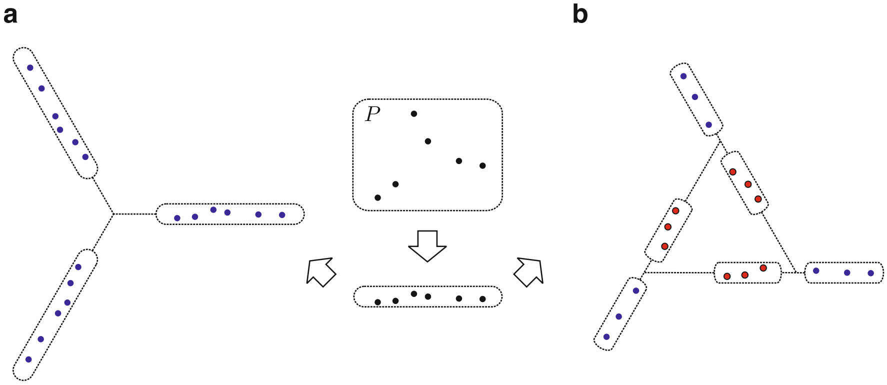


Fig. 1 (**a**) Recursive construction by Singer. (**b**) Recursive construction by Brodsky et al.

$$\overline{\text{cr}}(K_n) \leq 3\overline{\text{cr}}(K_{n/3}) + 3 \cdot \frac{n}{3}\binom{n/3}{3} + 3\binom{n/3}{2}^2.$$

Indeed, the first term consists of the crossings obtained from four points in the same wing, the next term counts the crossings from three points in one wing and the remaining in one of the other two wings, and the last term counts the crossings from two points in one wing and two points in another wing. Using $\overline{\text{cr}}(K_3) = 0$ as a starting point, this inequality gives $q_* \leq 5/13 < 0.38462$.

Brodsky et al. [18] modified Singer's construction by sliding three points in each wing toward the center of rotation as shown in Fig. 1b. Their construction gives $q_* \leq 6{,}467/16{,}848 < 0.38385$.

Aichholzer et al. [9] devised a different replacement construction. They started with an underlying set P with an even number of points N. Instead of triplicating P, they replaced every point of P by a cluster of c points on a small arc of circle flat enough so that all lines among these c points leave $N/2$ points of P on one side and $N/2 - 1$ on the other side (see Fig. 2a). Letting $n = cN$, their construction gives

$$\overline{\text{cr}}(K_n) \leq \left(\frac{24\overline{\text{cr}}(P) + 3N^3 - 7N^2 + 6N}{N^4}\right)\binom{n}{4} + \Theta(n^3).$$

Using a set P with $N = 36$ points and $\overline{\text{cr}}(P) = 21\,191$ they obtained $q_* < 0.380858$. They explored further using different sizes for each of the clusters, which resulted in an improvement of the latter bound to $q_* < 0.380739$. This method of obtaining lower bounds allowed for improvements by using better initial sets P. Aichholzer and Krasser [13], as part of their computer-assisted search of the crossing numbers $\overline{\text{cr}}(K_n)$ for small values of n, obtained a particular drawing of K_{54} that gives $q_* < 0.380601$.

Ábrego and Fernández-Merchant [5] started with an underlying set P with an even number of points N. They obtained a new set Q by replacing every point of P by a pair of points close to each other and spanning a line that divides the rest of Q

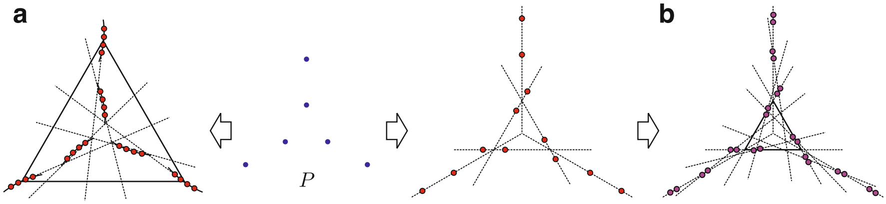


Fig. 2 (**a**) Replacement construction by Aichholzer et al. (**b**) Recursive construction by Ábrego and Fernández-Merchant

in half (see Fig. 2b). This property of having a *halving-line matching* is not satisfied by an arbitrary point set P, but fortunately it is satisfied by most of the small sets with optimal crossing number. Moreover, the resulting set Q inherits this property. Thus, if $n = 2^k N$, then iterating this construction k times gives

$$\overline{\mathrm{cr}}(K_n) \leq \left(\frac{24\overline{\mathrm{cr}}(P) + 3N^3 - 7N^2 + (30/7)N}{N^4} \right) \binom{n}{4} + \Theta(n^3). \tag{8}$$

At the time, using the best-known drawing of K_{30} (now proved to be optimal) yielded $q_* < 0.380559$. To this date, (8) provides the currently best recursive construction. The restrictions on the base set P were subsequently weakened [2] in the sense that (8) also holds for arbitrary sets P with an odd number of points. Applying this inequality to a drawing of K_{315} with 152,210,640 crossings gives the currently best upper bound: $q_* < \frac{83{,}247{,}328}{218{,}791{,}125} < 0.380488$.

To support the belief that the crossing-minimal sets have nearly threefold symmetry, Ábrego et al. [2] constructed a threefold symmetric set of n points for each n multiple of 3, $n < 100$ (see Fig. 3). Moreover, threefold symmetry is inherited from the base set in all recursive constructions mentioned before. In fact, the drawing of K_{315} used as a base set to obtain the best current upper bound has threefold symmetry.

7 Summary

In this section we summarize, for quick reference, the state of the art on $\overline{\mathrm{cr}}(K_n)$ and q_* at the time of writing this chapter.

7.1 Sylvester's Four-Point Constant

$$0.379972 < \frac{277}{729} \leq q_* \leq \frac{83{,}247{,}328}{218{,}791{,}125} < 0.380488. \tag{9}$$

The lower and upper bounds in (9) are derived in [2,3] (see also [7]), respectively.

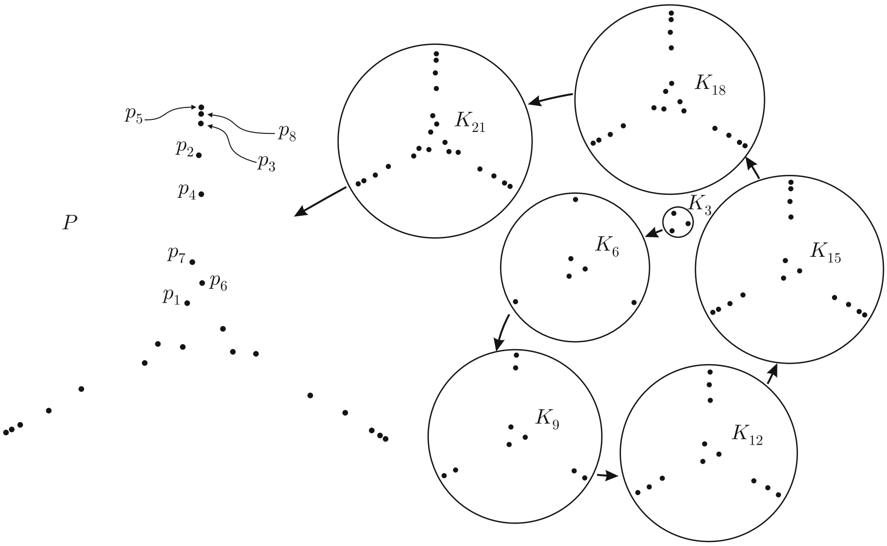


Fig. 3 The underlying vertex set of an optimal 3-symmetric geometric drawing of K_{24}. This point set contains optimal nested 3-symmetric drawings of $K_{21}, K_{18}, K_{15}, K_{12}, K_9, K_6$, and K_3

Table 1 Exact rectilinear crossing numbers known

n	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\overline{\text{cr}}(K_n)$	1	3	9	19	36	62	102	153	229	324	447	603	798	1,029

n	19	20	21	22	23	24	25	26	27	30
$\overline{\text{cr}}(K_n)$	1,318	1,657	2,055	2,528	3,077	3,699	4,430	5,250	6,180	9,726

7.2 Exact Values of $\overline{\text{cr}}(K_n)$

The exact value of $\overline{\text{cr}}(K_n)$ is known for $n \leq 27$ and for $n = 30$ (see Table 1).

For $n \leq 27$, the lower bound for $\overline{\text{cr}}(K_n)$ is derived in [3] (see also [6]). The bound $\overline{\text{cr}}(K_{30}) \geq 9726$ is proved in [19]. In all cases, the upper bounds were obtained by Aichholzer (http://www.ist.tugraz.at/staff/aichholzer/research/rp/triangulations/crossing/).

8 Further Thoughts and Future Research

Since the introduction of (2) in [4, 25], all the progress achieved on lower bounding q_* has been contingent on the derivation of improved bounds for $E_{\leq k}(n)$.

Although it may seem natural to expect the continuation of this trend, there is some evidence that suggests that this approach alone will not lead to the correct value of q_*. The reasons behind our caution lie in our own investigations of sets that minimize the number of ($\leq k$)-edges. So far it has been possible to construct n-point sets that simultaneously minimize $E_{\leq k}(n)$ for all k up to a certain value. It is not difficult to construct an n-point set that simultaneously achieves equality in (3) for every k, $0 \leq k \leq n/3$, and arbitrary n (along this discussion we assume that n is a multiple of 3). To construct a similar set minimizing $E_{\leq k}(n)$ for a larger range of values of k is notably harder. Aichholzer et al. [11] constructed an n-point set that simultaneously achieves equality in (6) for every k, $0 \leq k \leq \lfloor \frac{5n}{12} \rfloor - 1$ and arbitrary n. A different type of construction was used in [3] to simultaneously show that (6) is tight for every k, $0 \leq k \leq 4n/9 - 1$. However, this construction is far from crossing-optimal due to a dramatic increase in the number of ($\leq k$)-sets when $k \geq 4n/9$, and avoiding this seems impossible. That is, insisting on simultaneously minimizing $E_{\leq k}(n)$ for all k, for k as large as possible, seems to actually increase the crossing number of the point sets under consideration. In view of this, a new paradigm might be in order. It seems not only possible, but very likely, that the crossing-minimal drawings of K_n for large values of n are attained by point sets that are not even close to minimizing $E_{\leq k}$ for every $k < (4n/9) - 1$. A proper understanding of this intriguing behavior seems out of our reach at the present time.

Although (2) validates the efforts to lower bound $\overline{\text{cr}}(K_n)$ via lower bounding $E_{\leq k}(n)$, our previous remarks suggest that, no matter how good the estimates are, this may not suffice in order to determine q_*. It is quite conceivable that the (exact or asymptotic) value of $E_{\leq k}(n)$ be known for every k, and still the estimate for $\overline{\text{cr}}(K_n)$ obtained from plugging this into (6) does not correspond to the correct (at least asymptotic) value of $\overline{\text{cr}}(K_n)$.

The outlook from the upper bounds front is also unclear. We might all be just one clever idea away from a breakthrough construction that yields (at least asymptotically) crossing-minimal geometrical drawings of K_n. Using best-case heuristics, it can be shown that any recursive construction for large N, where each point is replaced by a small cluster of points of the same size, can yield at best a bound of the form

$$\overline{\text{cr}}(K_n) \leq \left(\frac{24\overline{\text{cr}}(P) + 3N^3 - 7N^2 + 4N}{N^4} \right) \binom{n}{4} + \Theta(n^3).$$

The improvement using the best drawing of K_{315} would be less than 10^{-8}.

For all these reasons, we are inclined to think that there is more potential to close the gap from below than from above; that is, we believe that q_* is closer to the current best upper bound than to the current best lower bound.

To end on an optimistic note, there is one more promising observation. Besides threefold symmetry, the currently best-known constructions (including those presented in [2]) share another property called 3-decomposability. A set P is called *3-decomposable* if there exists a balanced partition of P into three parts A, B, and C and a triangle T enclosing P such that the orthogonal projections of P onto the sides of T show A between B and C on one side, B between A and C on another side, and C between A and B on the third side. As for threefold symmetry, 3-decomposability is inherited from a base set in all recursive constructions mentioned before.

Ábrego et al. [2] conjectured that all crossing-minimal sets are 3-decomposable. If this conjecture happens to be true, then the lower bound for q_* would be improved to $(2/27)(15-\pi^2) > 0.380029$, as proved in [2].

References

1. B.M. Ábrego, J. Balogh, S. Fernández-Merchant, J. Leaños, G. Salazar, An extended lower bound on the number of $(\leq k)$-edges to generalized configurations of points and the pseudolinear crossing number of K_n. J. Comb. Theor. Ser. A **115**, 1257–1264 (2008)
2. B.M. Ábrego, M. Cetina, S. Fernández-Merchant, J. Leaños, G. Salazar, 3-Symmetric and 3-decomposable drawings of K_n. Discr. Appl. Math. **158**, 1240–1258 (2010)
3. B.M. Ábrego, M. Cetina, S. Fernández-Merchant, J. Leaños, G. Salazar, On $(\leq k)$-edges, crossings, and halving lines of geometric drawings of K_n. Discrete Comput. Geom. 48, 192–215 (2012)
4. B.M. Ábrego, S. Fernández-Merchant, A lower bound for the rectilinear crossing number. Graphs Comb. **21**, 293–300 (2005)
5. B.M. Ábrego, S. Fernández-Merchant, Geometric drawings of K_n with few crossings. J. Comb. Theor. Ser. A **114**, 373–379 (2007)
6. B.M. Ábrego, S. Fernández-Merchant, J. Leaños, G. Salazar, The maximum number of halving lines and the rectilinear crossing number of K_n for $n \leq 27$. Electron. Notes Discr. Math. **30**, 261–266 (2008)
7. B.M. Ábrego, S. Fernández-Merchant, J. Leaños, G. Salazar, A central approach to bound the number of crossings in a generalized configuration. Electron. Notes Discr. Math. **30**, 273–278 (2008)
8. O. Aichholzer, F. Aurehnammer, H. Krasser, Enumerating order types for small point sets with applications. Order **19**, 265–281 (2002)
9. O. Aichholzer, F. Aurenhammer, H. Krasser, On the crossing number of complete graphs. Computing **76**, 165–176 (2006)
10. O. Aichholzer, J. García, D. Orden, P. Ramos, New lower bounds for the number of $(\leq k)$-edges and the rectilinear crossing number of K_n. Discr. Comput. Geom. **38**, 1–14 (2007)
11. O. Aichholzer, J. García, D. Orden, P. Ramos, New results on lower bounds for the number of $(\leq k)$-facets. Eur. J. Comb. **30**, 1568–1574 (2009)
12. O. Aichholzer, H. Krasser, The point set order type data base: a collection of applications and results. in *Proceedings of the 13th Annual Canadian Conference on Computational Geometry CCCG 2001*, Waterloo, Ontario, 2001, pp. 17–20
13. O. Aichholzer, H. Krasser, Abstract order type extension and new results on the rectilinear crossing number. Comput. Geom. Theor. Appl. **36**, 2–15 (2006)
14. J. Balogh, G. Salazar, k-Sets, convex quadrilaterals, and the rectilinear crossing number of K_n. Discr. Comput. Geom. **35**, 671–690 (2006)
15. W. Blaschke, Über Affine Geometrie XI: Lösung des "Vierpunktproblems" von Sylvester aus der Theorie der geometrischen Wahrscheinlichkeiten. Leipziger Berichte **69**, 436–453 (1917)
16. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, New York, 2005)
17. A. Brodsky, S. Durocher, E. Gethner, The rectilinear crossing number of K_{10} is 62. Electron. J. Comb. **8**, R23 (2001)
18. A. Brodsky, S. Durocher, E. Gethner, Toward the rectilinear crossing number of K_n: new drawings, upper bounds, and asymptotics. Discr. Math. **262**, 59–77 (2003)

19. M. Cetina, C. Hernández-Vélez, J. Leaños, C. Villalobos, Point sets that minimize $(\leq k)$-edges, 3-decomposable drawings, and the rectilinear crossing number of K_{30}. Discr. Math. **311**, 1646–1657 (2011)
20. M.W. Crofton, Probability, in *Encyclopedia Britannica*, 9th edn. Philadelphia, PA: J. M. Stoddart **19**, 768–788 (1885)
21. R.B. Eggleton, Ph.D. thesis, University of Calgary, 1973
22. R.K. Guy, Crossing numbers of graphs. Graph theory and applications, in *Proceeding Conference*, Western Michigan University, Kalamazoo, Mich., dedicated to the memory of J. W. T. Youngs), Lecture Notes in Math., Springer, Berlin, **303**, 111–124 (1972)
23. H.F. Jensen, An upper bound for the rectilinear crossing number of the complete graph. J. Combin. Theor. Ser B **10**, 212–216 (1971)
24. H. Krasser, Order Types of Point Sets in the Plane. Ph.D. thesis, Institute for Theoretical Computer Science, Graz University of Technology, Austria, October 2003
25. L. Lovász, K. Vesztergombi, U. Wagner, E. Welzl, Convex quadrilaterals and k-sets. Towards a theory of geometric graphs. Contemp. Math. (Amer. Math. Soc., Providence) **342**, 139–148 (2004)
26. S. Pan, R.B. Richter, The crossing number of K_{11} is 100. J. Graph Theor. **56**, 128–134 (2007)
27. E. Scheinerman, H.S. Wilf, The rectilinear crossing number of a complete graph and Sylvester's "four point" problem of geometric probability. Am. Math. Mon. **101**, 939–943 (1994)
28. D. Singer, Rectilinear crossing numbers. Manuscript (1971)
29. J.J. Sylvester, Question 1491, in *The Educational Times* (London, April 1864)
30. J.J. Sylvester, Rep. Br. Assoc. **35**, 8–9 (1865)
31. U. Wagner, On the rectilinear crossing number of complete graphs, in *Proceedings of the 14th Annual Symposium on Discrete Algorithms* (Baltimore, MD, 2003), ACM, New York, 583–588 (2003)
32. U. Wagner, E. Welzl, A continuous analogue of the upper bound theorem. Discr. Comput. Geom. **26**, 205–219 (2001)
33. E. Welzl, More on k-sets of finite sets in the plane. Discr. Comput. Geom. **1**, 95–100 (1986)
34. E. Welzl, Entering and leaving j-facets. Discr. Comput. Geom. **25**, 351–364 (2001)

The Maximum Number of Tangencies Among Convex Regions with a Triangle-Free Intersection Graph

Eyal Ackerman

Abstract Let $t(\mathcal{C})$ be the number of tangent pairs among a set $\mathcal{C}$ of n Jordan regions in the plane. Pach et al. [Tangencies between families of disjoint regions in the plane, in *Proceedings of the 26th ACM Symposium on Computational Geometry (SoCG 2010)*, Snowbird, June 2010, pp. 423–428] showed that if $\mathcal{C}$ consists of convex bodies and its *intersection graph* is bipartite, then $t(\mathcal{C}) \leq 4n - \Theta(1)$, and, moreover, there are such sets that admit at least $3n - \Theta(\sqrt{n})$ tangencies. We close this gap and generalize their result by proving that the correct bound is $3n - \Theta(1)$, even when the intersection graph of $\mathcal{C}$ is only assumed to be triangle-free.

1 Introduction

Erdős's famous *unit distance* problem [3, 4] asks for the maximum number of pairs of points that are at unit distance from each other among n distinct points in the plane. This is equivalent to asking for the maximum number of *tangency points* among n distinct unit disks in the plane.

Let $\mathcal{C}$ be a family of Jordan regions in the plane. We say that two regions are *tangent* if they intersect at a single point, and denote by $t(\mathcal{C})$ the number of tangent pairs in $\mathcal{C}$. It is easy to see that n (convex) regions might determine $\Theta\left(n^2\right)$ tangency points: For instance, there are n^2 tangency points (*tangencies*) between the $2n$ regions consisting of the n sides of a convex n-gon and a set of n convex n-gons each of which has a vertex on each of the sides of the first polygon. However, more restricted families of regions might determine fewer tangencies. One example are unit disks (Erdős's unit distance problem): They are are known to admit at

E. Ackerman (✉)
Department of Mathematics, Physics, and Computer Science,
University of Haifa at Oranim, Tivon 36006, Israel
e-mail: ackerman@sci.haifa.ac.il

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_3,

most $O\left(n^{4/3}\right)$ tangencies [6], and it is conjectured that the correct bound is Erdős's lower bound of $\Omega\left(n\,\mathrm{e}^{\frac{c\log n}{\log\log n}}\right)$ [4] (see also [3] for the history of this problem and additional references). Another example is regions such that the boundary curves of every pair of them intersect exactly once or twice, and no three boundary curves intersect at a point. Such regions admit only $O(n)$ tangencies [1, 2].

For a set of regions $\mathcal{C}$, denote by $\mathcal{I}(\mathcal{C})$ the *intersection graph* of $\mathcal{C}$, that is, the graph whose vertex set is $\mathcal{C}$ and whose edge set consists of all the intersecting pairs. Ben-Dan and Pinchasi (2007, personal communication) observed that if $\mathcal{I}(\mathcal{C})$ is bipartite, then $t(\mathcal{C}) = O\left(n^{3/2}\log n\right)$, and they suggested that the correct bound is $O(n)$. Pach et al. [5] proved this conjecture for the case of convex regions and found almost matching lower and upper bounds for the maximum number of tangencies.

Theorem 1 ([5]). *Let $\mathcal{C}$ be a set of n convex bodies in the plane. If $\mathcal{I}(\mathcal{C})$ is bipartite, then $t(\mathcal{C}) \le 4n - \Theta(1)$. Moreover, for every n, there is set $\mathcal{C}$ of n convex bodies in the plane such that $\mathcal{I}(\mathcal{C})$ is bipartite and $t(\mathcal{C}) \ge 3n - \Theta(\sqrt{n})$.*

They also suggested the following conjecture.

Conjecture 2 ([5]). For every fixed integer $k > 2$, if $\mathcal{C}$ is a set of n convex bodies in the plane such that $\mathcal{I}(\mathcal{C})$ is K_k-free, then $t(\mathcal{C}) \le c_k n$, for some constant c_k that depends only on k.

In this note we close the gap in Theorem 1 and prove Conjecture 2 for $k = 3$.

Theorem 3. *Let $\mathcal{C}$ be a set of n convex bodies in the plane. If $\mathcal{I}(\mathcal{C})$ is triangle-free, then $t(\mathcal{C}) \le 3n - \Theta(1)$. Moreover, for every n, there is set $\mathcal{C}$ of n convex bodies in the plane such that $\mathcal{I}(\mathcal{C})$ is bipartite and $t(\mathcal{C}) \ge 3n - \Theta(1)$.*

2 Proof of Theorem 3

Most of the proof is devoted to the upper bound. For the lower bound, see Proposition 2.13.

Let $\mathcal{C}$ be a set of $n \ge 4$ convex bodies in the plane. We prove by induction on n that if $\mathcal{I}(\mathcal{C})$ is triangle-free, then $t(\mathcal{C}) \le 3n - 6$. The first steps of the proof follow the proof of [5, Theorem 7]. Since $t(\mathcal{C}) \le \binom{n}{2}$, for $n = 4$ there are at most $\binom{4}{2} = 6 \le 3n - 6$ tangencies. Suppose now that $n \ge 5$ and that the theorem holds for every $\mathcal{C}'$ as above, with $4 \le |\mathcal{C}'| < n$. Let $\mathcal{C}$ be a set of n convex bodies in the plane, such that $\mathcal{I}(\mathcal{C})$ is triangle-free; that is, $\mathcal{C}$ does not contain three pairwise intersecting bodies. We may assume that every body in $\mathcal{C}$ is tangent to at least four other bodies, for otherwise we can conclude by induction. We begin by replacing every convex body $C \in \mathcal{C}$ by a convex polygon whose vertices are the tangency points along the boundary of C. Henceforth, $\mathcal{C}$ denotes the set of convex polygons.

Proposition 2.1. *There are no two polygons $P, Q \in \mathcal{C}$ such that a vertex of P is inside Q.*

Proof. Suppose there is such a vertex v. Then P touches another polygon $R \neq P, Q$ at v. Therefore, P, Q, R are pairwise intersecting. □

Denote by T the set of tangency points, and let $m = |T|$. Next, we define a planar graph $G = (V, E)$ by placing a vertex at every point in T and at every intersection point between sides of two polygons. Note that this graph is 4-regular. Denote by F the set of faces of G, and by $|f|$ the size of a face $f \in F$. A *k-face* is a face of size k. We write $f \subseteq P$ when a face f is contained in a polygon P. Note that each face $f \in F$ is contained in exactly 0, 1, or 2 polygons, since $\mathcal{I}(\mathcal{C})$ is triangle-free. Denote by F_i the set of faces that are contained in exactly i polygons, for $i = 0, 1, 2$. We proceed with a few observations on G, most of them already appear in [5].

Proposition 2.2. *Let v be a vertex of a face $f \in F_1$. If $v \notin T$, then one of its neighbors in f is also a crossing point.*

Proof. Suppose that $v \notin T$ is a vertex of $f \in F_1$ and let $P \in \mathcal{C}$ be the polygon that contains f. Then v is the intersection point of a side of P and a side of another polygon Q. This side must intersect P at another point u, since by Proposition 2.1, P does not contain a vertex of Q. The segment vu cannot be crossed by a side of another polygon, since there are no three pairwise intersecting polygons. Therefore, uv is an edge of f, and $u \notin T$ is a neighbor of v in f. □

Proposition 2.3. *If $f \in F_0 \cup F_1$ is a 3-face, then $f \in F_1$ and f has exactly one vertex from T.*

Proof. The three edges of f must be contained in sides of two polygons. Indeed, if all of them are contained in sides of one polygon, then this polygon is a triangle; however, we assumed that any polygon has at least four vertices. Otherwise, if each edge of f is contained in a side of a different polygon, then we have three pairwise intersecting polygons.

Therefore, f has two edges that are contained in sides of the same polygon; hence, they intersect at a vertex of this polygon and $f \in F_1$. The third edge of f must belong to a side of another polygon; thus, the remaining vertices of f are crossing points. □

Every tangency point $t \in T$ is adjacent to two faces from F_1 and to two faces from F_0. Define $F(t) \stackrel{\text{def}}{=} (|f_1|, |f_2|, |f_3|, |f_4|)$, where the faces f_i, $1 \leq i \leq 4$, are the four faces adjacent to t, such that $f_1, f_2 \in F_1$, $|f_1| \leq |f_2|$, and $f_3, f_4 \in F_0$, $|f_3| \leq |f_4|$. We may assume that all the faces adjacent to a tangency point are distinct, for otherwise G has a cut vertex and we can conclude by induction. Call a vertex *bad* if it is adjacent to at least one 3-face, and *double* bad if it is adjacent to two 3-faces. The next observation follows from Proposition 2.3.

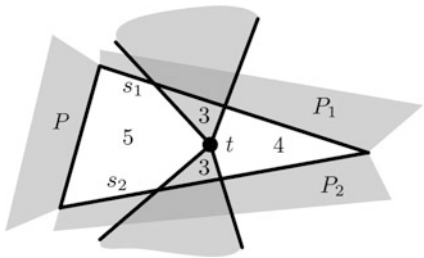


Fig. 1 $F(t) = (3,3,4,5)$ yields three pairwise intersecting polygons

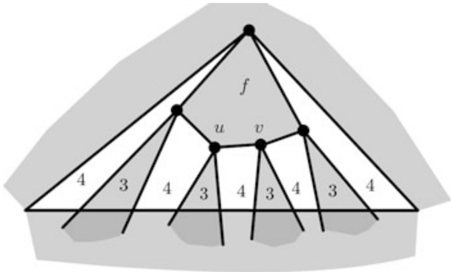


Fig. 2 If f has $|f|-1$ $(3,|f|,4,4)$-vertices on its boundary, then the remaining vertex is concave

Observation 2.4. *If $t \in T$ is a bad vertex, then in each of the two faces from F_0 that are adjacent to t, at least one neighbor of t is a crossing point. If t is double bad, then all of its neighbors are crossing points.*

Proposition 2.5. *If $t \in T$ is a double bad vertex, then $F(t) \in \{(3,3,4,\geq 6),(3,3,\geq 5,\geq 5)\}$.*

Proof. Let $F(t) = (|f_1|,|f_2|,|f_3|,|f_4|)$, such that $|f_1| = |f_2| = 3$. It follows from Proposition 2.3 that $|f_3| \geq 4$. Since $|f_3| \leq |f_4|$, then clearly if $|f_3| \geq 5$, then $|f_4| \geq 5$. Suppose that $|f_3| = 4$. For $i = 1,2$, let e_i be the opposite edge to t in f_i, and let s_i be the side of the polygon P_i that contains e_i. Since $|f_3| = 4$, the vertex opposite t in f_3 must be an intersection point of s_1 and s_2. Clearly, s_1 and s_2 intersect once, so $|f_4| \geq 5$. However, if $|f_4| = 5$, then there must be a side of a polygon $P \neq P_1, P_2$ that intersects s_1 and s_2; hence, we have three pairwise intersecting polygons (see Fig. 1). Therefore, $|f_4| \geq 6$ in this case. □

Proposition 2.6. *Any face $f \in F_1$ has at most $|f|-2$ vertices $t \in T$ on its boundary such that $F(t) = (3,|f|,4,4)$.*

Proof. Let f be a face in F_1, and suppose there are two neighboring vertices on its boundary u,v such that $F(u) = F(v) = (3,|f|,4,4)$. Then there is one polygon side that supports the 3-faces adjacent to u or v and the 4-faces from F_0 that are adjacent to u or v (see Fig. 2). Therefore, if f has $|f|$ $(3,|f|,4,4)$-vertices on its boundary, then there is a polygon side that surrounds f. If f has exactly $|f|-1$ such vertices, then the remaining vertex must be a concave vertex of a polygon (see Fig. 2). □

We proceed by assigning every face $f \in F$ a *weight* $w(f) = |f| - 4$. Let W be the total weight we assigned, and observe that by Euler's polyhedral formula we have

$$W = \sum_{f\in F}(|f|-4) = \sum_{f\in F}(|f|-4) + \sum_{v\in V}(\deg(v)-4) = 4(|E|-|F|-|V|) = -8. \quad (1)$$

Proposition 2.7. *For every polygon $P \in \mathcal{C}$, it holds that $w(P) \stackrel{\text{def}}{=} \sum_{f\subseteq P} w(f) = |P| - 4$.*

Proof. Since there are no three pairwise intersecting polygons, the interior of P is divided into faces by disjoint segments connecting pairs of interior points on the sides of P. Assume that we add these segments one by one, while keeping track of $w(P)$. Every new segment we add increases the number of faces by one [thus contributing -4 to $w(P)$] and increases by 4 the number of sides of faces in P. Therefore, $w(P)$ maintains its initial value, which is $|P| - 4$. □

Every tangency point is a vertex of exactly two polygons. Therefore,

$$\sum_{P\in\mathcal{C}}(|P|-4) = 2m - 4n. \quad (2)$$

Combining (1) and (2), we get

$$\begin{aligned} -8 = W &= \sum_{f\in F_1} w(f) + \sum_{f\in F_2} w(f) + \sum_{f\in F_0} w(f) \\ &= \frac{1}{2}\sum_{P\in\mathcal{C}} w(P) + \frac{1}{2}\sum_{f\in F_1} w(f) + \sum_{f\in F_0} w(f) \\ &= m - 2n + \frac{1}{2}\sum_{f\in F_1} w(f) + \sum_{f\in F_0} w(f). \end{aligned} \quad (3)$$

Pach et al. [5] showed that if $\mathcal{I}(\mathcal{C})$ is bipartite, then

$$\frac{1}{2}\sum_{f\in F_1} w(f) + \sum_{f\in F_0} w(f) \geq -m/2, \quad (4)$$

which, when plugged into (3), gives $m \leq 4n - 16$. We use the *discharging method* to refine their analysis and replace the right-hand side of (4) by $-m/3$, and obtain $m \leq 3n - 12 \leq 3n - 6$. Namely, we prove

Lemma 2.8. $\frac{2}{3}m + \sum_{f\in F_1}(|f|-4) + 2\sum_{f\in F_0}(|f|-4) \geq 0.$

Proof. We assign initial *charges* as follows.

- For every face $f \in F_1$, we set $\text{ch}_0(f) = w(f) = |f| - 4$.
- For every face $f \in F_0$, we set $\text{ch}_0(f) = 2w(f) = 2|f| - 8$.
- For every tangency point $t \in T$, we set $\text{ch}_0(t) = \frac{2}{3}$.

It remains to show that the total initial charge is nonnegative. We do that by redistributing the charges (*discharging*) in several rounds, such that the total charge remains the same, and eventually, every element has a nonnegative final charge. We denote by $\mathrm{ch}_i(x)$ the charge of an element x after the ith discharging round. Note that the only elements with a negative initial charge are 3-faces, whose charge is -1. A vertex is *good* if it has a positive charge.

Round 1. A face $f \in F_0$ such that $|f| \in \{6,7\}$ sends $\frac{4}{3}$ units of charge (henceforth, "units") to each double bad vertex on its boundary and distributes the rest of its charge evenly to every other tangency point on its boundary. Any other face $f' \in F_0 \cup F_1$ sends $\mathrm{ch}_0(f')/k$ units to each of the k tangency points on its boundary.

Proposition 2.9. *Let $f \in F_0$ be a face and let t be a tangency point on its boundary. Then the following holds in Round 1:*

(i) If t is double bad and $|f| = 5$, then f sends at least $\frac{2}{3}$ units to t.
(ii) If t is double bad and $|f| > 5$, then f sends at least $\frac{4}{3}$ units to t.
(iii) If $|f| = 5$, then f sends at least $\frac{2}{5}$ units to t, and at least $\frac{2}{4}$ units if t is also adjacent to a 3-face.
(iv) If $|f| \geq 6$, then f sends at least $\frac{2}{3}$ units to t.

Proof. The claims follow from the definition of Round 1 and from Observation 2.4. □

Proposition 2.10. *After Round 1, the following holds:*

(i) Every face in $F_0 \cup F_1$ has a nonnegative charge.
(ii) For every vertex $t \in T$ if $\mathrm{ch}_1(t) < 0$, then $\mathrm{ch}_1(t) = -\frac{1}{3}$ or $\mathrm{ch}_1(t) = -\frac{2}{15}$. In the first case $F(t) = (3,4,4,4)$, while in the second case $F(t) = (3,5,4,4)$.

Proof. Observe that by Proposition 2.3, every 3-face has one vertex from T to which it sends its negative charge and ends up with charge zero. Every face f has at most $\lfloor |f|/2 \rfloor$ double bad vertices on its boundary; therefore, 6- and 7-faces from F_0 remain with a nonnegative charge. Any other face clearly remains with a nonnegative charge, and therefore (i) holds.

For the second claim, note that if $\mathrm{ch}_1(t) < 0$, then t must be a bad vertex. If t is double bad, then it follows from Proposition 2.5 that $F(t) = (3,3,4,\geq 6)$ or $F(t) = (3,3,\geq 5,\geq 5)$. By Proposition 2.9, in the first case $\mathrm{ch}_1(t) \geq \frac{2}{3} + 2 \cdot (-1) + \frac{4}{3} \geq 0$, while in the second case, $\mathrm{ch}_1(t) \geq \frac{2}{3} + 2 \cdot (-1) + 2 \cdot \frac{2}{3} \geq 0$.

Finally, suppose that t is adjacent to exactly one 3-face. If t is adjacent to a face f such that f is a 6-face or $|f| = 5$ and $f \in F_0$, then t receives from f at least $\frac{1}{3}$ units and ends up with a nonnegative charge. The other cases are listed (note that by Proposition 2.2, a 5-face in F_1 sends either $\frac{1}{5}$ or at least $\frac{1}{3}$ units to each tangency point on its boundary). □

Round 2. Let $t \in T$ be a tangency point with $\mathrm{ch}_1(t) < 0$. Then t is adjacent to a 3-face and therefore has at most two good neighbors. If t has exactly one good neighbor or $\mathrm{ch}_1(t) > -\frac{2}{3}$, then t asks for $-\mathrm{ch}_1(t)$ units from one of its good

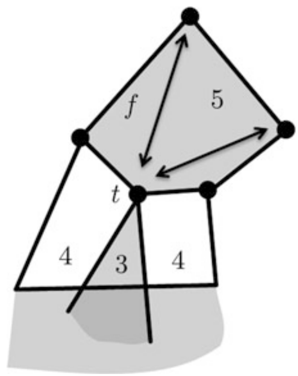


Fig. 3 Round 4: If $F(t) = (3,5,4,4)$ and $\mathrm{ch}_3(t) = -\frac{2}{15}$, then t ask for $\frac{1}{15}$ unit from each of its nonneighbors in f

neighbors. If t has two good neighbors and $\mathrm{ch}_1(t) = -\frac{2}{3}$, then t asks for $\frac{1}{6}$ units from each of its good neighbors. If a good vertex $q \in T$ is being asked for ε units by a vertex t, then q accepts t's request and sends it ε units if and only if $\varepsilon \leq \mathrm{ch}_1(q)/j$, where j is the number of requests q got.

Round 3. Repeat Round 2 with respect to $\mathrm{ch}_2(\cdot)$.

Round 4. Suppose that $F(t) = (3,5,4,4)$ and let f be the 5-face that is adjacent to t. If all the vertices of f are from T and $\mathrm{ch}_3(t) = -\frac{2}{15}$, then t asks for $\frac{1}{15}$ units from each of the two vertices of f that are not its neighbors (see Fig. 3). They accept t's requests if they can afford it (as in Round 2).

Clearly, for any $t \in T$ and $i \geq 1$, if $\mathrm{ch}_i(t) \geq 0$, then $\mathrm{ch}_{i+1}(t) \geq 0$. Thus, by Proposition 2.10 it remains to verify that $\mathrm{ch}_4(t) \geq 0$ for $t \in T$ such that $F(t) \in \{(3,4,4,4),(3,5,4,4)\}$.

Proposition 2.11. *If* $F(t) = (3,4,4,4)$, *then* $\mathrm{ch}_3(t) \geq 0$.

Proof. Let f_1 be the 4-face from F_1 that is adjacent to t, let p and q be the vertices adjacent to t in f_1, and let r be the vertex opposite to t in f_1. If neither p nor q is in T, then the local neighborhood of t looks like Fig. 4a. If r is a crossing point, then we have three pairwise intersecting polygons. Otherwise, if $r \in T$, then it is a concave vertex of a polygon. Therefore, we may proceed by considering the case in which $p, q \in T$ and the case that exactly one of p and q is in T. In the latter case we assume, without loss of generality, that $p \in T$ and $q \notin T$.

Case 1: $p \in T$ and $q \notin T$. Since p is the only good neighbor of t, t asks $\frac{1}{3}$ units from p in Round 2 (and again in the next round if its first request is denied). Note that in this case $r \notin T$ by Proposition 2.2, and therefore t is the only bad neighbor of p. Let f_2 be the other face from F_1 that is adjacent to p. We consider four subcases, according to whether $|f_2| = 3, 4, 5$ or $|f_2| \geq 6$.

Subcase 1a: $|f_2| = 3$**.** Refer to Fig. 4b and consider $|f_3|$. If $|f_3| = 4$, then the segments s_1 and s_2 must intersect twice. If $|f_3| = 5$, then there are three pairwise intersecting polygons. Therefore, $|f_3| \geq 6$ and (by Proposition 2.9) it sends at least

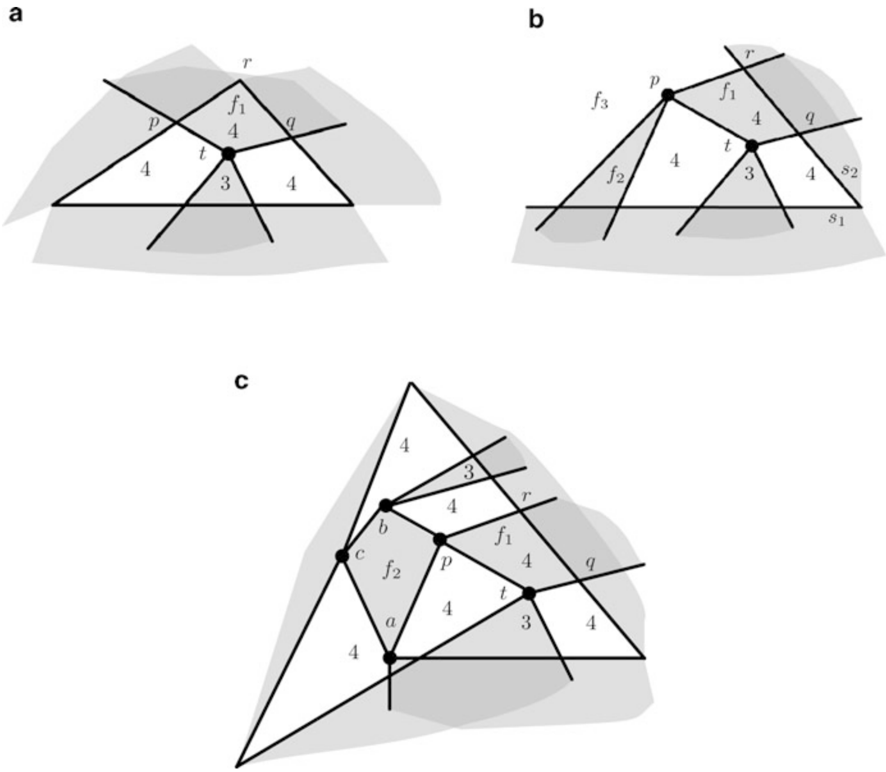


Fig. 4 Case 1 in the proof of Proposition 2.11

$\frac{2}{3}$ units to p in Round 1. Thus, $\text{ch}_1(p) \geq \frac{2}{3} - 1 + \frac{2}{3} = \frac{1}{3}$. Since t is the only bad neighbor of p, p accepts t's request in Round 2, and $\text{ch}_2(t) = 0$.

Subcase 1b: $|f_2| = 4$. Since $F(p) = (4,4,4,\geq 4)$, we have $\text{ch}_1(p) \geq \frac{2}{3}$. If p has at most one other neighbor but t that asks for charge in Round 2, then p can accept t's request in this round and $\text{ch}_2(t) = 0$. Otherwise, p has exactly three neighbors with a negative charge after Round 1; refer to Fig. 4c. Since both a and b are adjacent to f_2, they must each be adjacent to a 3-face, and to two 4-faces from F_0. That is, $F(a) = F(b) = (3,4,4,4)$. Observe that $F(c) \neq (3,4,4,4)$, by Proposition 2.6. Therefore, $\text{ch}_1(c) \geq \frac{2}{3}$, and in Round 2 each of a and b asks (and receives) only $\frac{1}{6}$ units from each of p and c. Thus, $\text{ch}_2(a), \text{ch}_2(b) = 0$ and in the next round p can accept t's request.

Subcase 1c: $|f_2| = 5$. In this case $\text{ch}_1(p) \geq \frac{13}{15}$. If p has at most one other neighbor but t that requests charge in Round 2, then p can accept t's request in this round and $\text{ch}_2(t) = 0$. Otherwise, p has exactly three bad neighbors asking for charge in Round 2. Denote by a and b the other two, and observe that both of them are adjacent to f_2. Therefore, $F(a) = F(b) = (3,5,4,4)$, and $\text{ch}_1(a) = \text{ch}_1(b) = -\frac{2}{15}$.

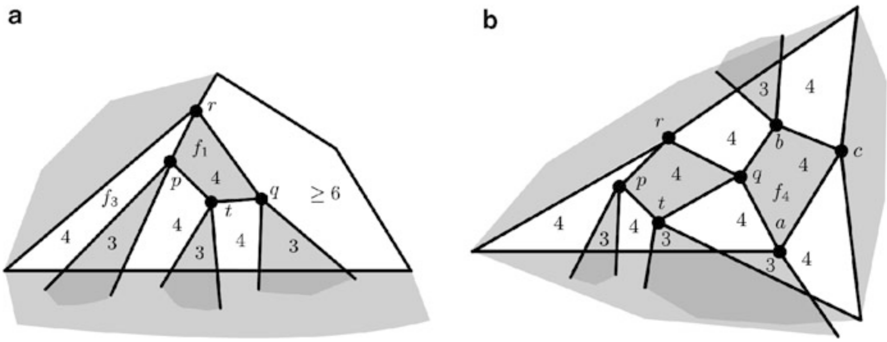


Fig. 5 Subcase 2a: $p, q \in T$, $F(t) = F(p) = (3,4,4,4)$

If follows that in Round 2 p sends at most $\frac{2}{15}$ units to each of them and $\frac{1}{3}$ units to t in the next round.

Subcase 1d: $|f_2| \geq 6$**.** In this case t is the only neighbor of p with a negative charge at the beginning of Round 2, and so p can accept t's request.

Case 2: $p \in T$ *and* $q \in T$. Note that in this case $r \in T$ as well. The first subcase that we consider is when $\text{ch}_1(p) < 0$. By symmetry the same arguments apply when $\text{ch}_1(q) < 0$, so the second subcase we need to consider is $\text{ch}_1(p), \text{ch}_1(q) \geq 0$.

Subcase 2a: $\text{ch}_1(p) < 0$**, that is,** $F(p) = (3,4,4,4)$**.** Since $\text{ch}_1(p) < 0$, in Round 2 (and perhaps also in Round 3) t might only ask $\frac{1}{3}$ units from q (if q has a positive charge). Observe that $\text{ch}_1(q), \text{ch}_1(r) \geq 0$. Indeed, if $\text{ch}_1(q) < 0$, then $F(q) = (3,4,4,4)$. But then f_1 has three $(3,4,4,4)$-vertices, which contradicts Proposition 2.6. For the same reason, $\text{ch}_1(r) \geq 0$.

Observe that $F(q) \neq (3,4,4,5)$, for otherwise there must be three pairwise intersecting polygons (refer to Fig. 5a). If q is not adjacent to a 3-face or $F(q) = (3,4,4,>6)$, then $\text{ch}_1(q) \geq \frac{2}{3}$. Therefore, if $\text{ch}_1(q) < \frac{2}{3}$, then $F(q) = (3,4,4,6)$, in which case $\text{ch}_1(q) \geq \frac{1}{3}$. We now consider both possibilities.

Subcase 2a.i: $F(q) = (3,4,4,6)$**.** Refer to Fig. 5a. Since q is adjacent to a 3-face, its only neighbors from T are t and r. Because $\text{ch}_1(r) \geq 0$, t in the only neighbor of q that requests charge at Round 2, so q accepts t's request and $\text{ch}_2(t) = 0$.

Subcase 2a.ii: $\text{ch}_1(q) \geq \frac{2}{3}$**.** If q has at most two neighbors asking charge in Round 2, then q can send $\frac{1}{3}$ units to t in Round 2. Otherwise, let a and b be the other two neighbors of q that ask charge in Round 2, and let f_4 be the face that is adjacent to a, b, and q (see Fig. 5b). Since $\text{ch}_1(a), \text{ch}_1(b) < 0$, either $|f_4| = 4$ or $|f_4| = 5$. We consider these two possibilities.

- Suppose that $|f_4| = 4$ and let c denote its remaining vertex. Then $c \in T$, and observe that $\text{ch}_1(c) > 0$. Indeed, if $\text{ch}_1(c) < 0$, then $F(c) = (3,4,4,4)$, which contradicts Proposition 2.6. Therefore, each of a and b asks only $\frac{1}{6}$ units from q at Round 2, and q is able to accept t's request at the next round.

- Suppose that $|f_4| = 5$. In this case we have $\text{ch}_1(q) \geq \frac{13}{15}$ and $\text{ch}_1(a), \text{ch}_1(b) \geq -\frac{2}{15}$. Therefore, q accepts t's request in Round 3.

Subcase 2b: $\text{ch}_1(p), \text{ch}_1(q) \geq 0$**.** Observe that in this case $\text{ch}_1(p), \text{ch}_1(q) \geq \frac{1}{6}$. Indeed, if p is not adjacent to a 3-face, then $\text{ch}_1(p) \geq \frac{2}{3}$. Otherwise, $F(p) = (3,4,4,\geq 5)$ and therefore $\text{ch}_1(p) \geq \frac{2}{3} - 1 + \frac{2}{4} = \frac{1}{6}$. For the same reason $\text{ch}_1(q) \geq \frac{1}{6}$. Therefore, t asks in Round 2 for $\frac{1}{6}$ units from each of p and q. Note that t's requests are accepted since, as we already observed, if p is not adjacent to a 3-face, then $\text{ch}_1(p) \geq \frac{2}{3}$ and, thus, it can accept t's request since it has at most four asking neighbors. Otherwise, $\text{ch}_1(p) \geq \frac{1}{6}$ and t is the only bad neighbor of p (since r is adjacent to a face from F_0 of size at least 5). Therefore, both p and q accept t's request in Round 2.
This concludes the proof of Proposition 2.11. □

Proposition 2.12. *If* $F(t) = (3,5,4,4)$*, then* $\text{ch}_4(t) \geq 0$.

Proof. Let $f_1 \in F_1$ be the 5-face adjacent to t. If f_1 has at most three tangency points, then each of them gets at least $\frac{1}{3}$ units in Round 1, and so $\text{ch}_2(t) \geq 0$. It remains to consider the case where all the vertices of f_1 are from T; i.e., f_1 is a polygon from $\mathcal{C}$. (By Proposition 2.2, f_1 cannot have exactly one vertex that is a crossing point.) Let p and q be the neighbors of t in f_1. If at least one of them has a positive charge after Round 1, then t asks one of them for $\frac{2}{15}$ units in Round 2. The other case to consider is when $\text{ch}_1(p), \text{ch}_1(q) < 0$.

Case 1: $\text{ch}_1(p) > 0$ *or* $\text{ch}_1(q) > 0$. Assume, without loss of generality, that $\text{ch}_1(p) > 0$ and t asks p for $\frac{2}{15}$ units in Round 2. If p is not adjacent to a 3-face, then $\text{ch}_1(p) \geq \frac{13}{15}$. In this case, in Round 2 p accepts t's request, since p may send at least $\frac{13}{15}/4 > \frac{2}{15}$ to each of its at most four bad neighbors. Otherwise, since p receives $\frac{1}{5}$ units from f_1 in Round 1 and $\text{ch}_1(p) > 0$, it follows that p is adjacent to a face of size at least 5 from F_0. That is, $F(p) = (3,5,4,\geq 5)$. By Proposition 2.9, this face sends at least $\frac{2}{4}$ units to p in Round 1; therefore, $\text{ch}_1(p) \geq \frac{2}{3} - 1 + \frac{1}{5} + \frac{2}{4} = \frac{11}{30}$. Note that t is the only bad neighbor of p, and thus p accepts t's request in Round 2.

Case 2: $\text{ch}_1(p), \text{ch}_1(q) < 0$. Then by Proposition 2.10, $F(p) = F(q) = (3,5,4,4)$. Let $a, b \in T$ be the other vertices of f_1 (see Fig. 6).

It follows from Proposition 2.6 that $F(a), F(b) \neq (3,5,4,4)$; therefore, $\text{ch}_1(a), \text{ch}_1(b) \geq 0$. Moreover, we claim that $\text{ch}_2(a) \geq \frac{k}{15}$, where $k \in \{1,2\}$ is the number of 5-faces from F_1 that are adjacent to a. Indeed, if $F(a) = (3,5,4,\geq 5)$, then $\text{ch}_1(a) \geq \frac{11}{30}$, and only p requests charge ($\frac{2}{15}$ units) from a in Round 2. Hence, $\text{ch}_2(a) \geq \frac{7}{30} > \frac{1}{15}$. Otherwise, if a is not adjacent to a 3-face, then $\text{ch}_1(a) \geq \frac{2}{3} + \frac{k}{5}$ and $\text{ch}_2(a) \geq \frac{2}{3} + \frac{k}{5} - 2 \cdot \frac{1}{3} - \frac{2}{15} = \frac{3k-2}{15} \geq \frac{k}{15}$.

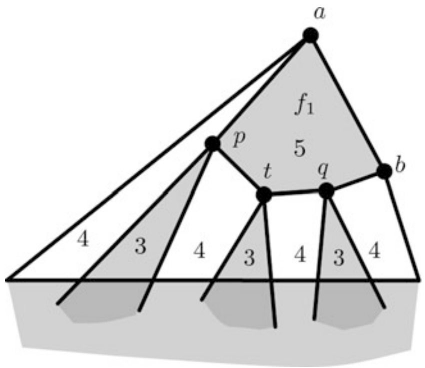


Fig. 6 $F(t) = F(p) = F(q) = (3,5,4,4)$

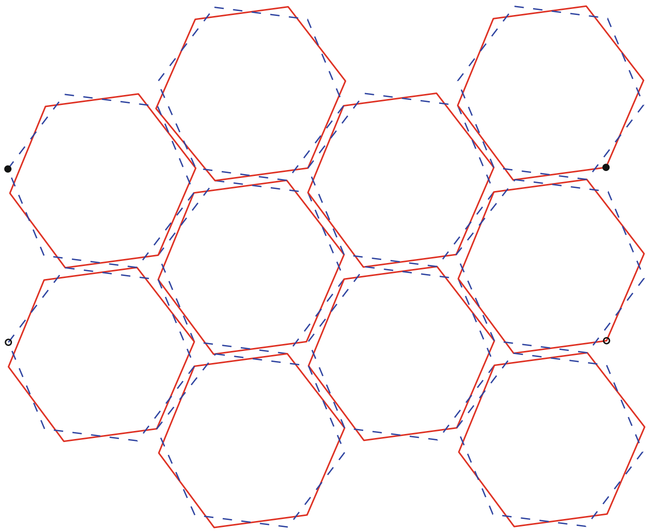


Fig. 7 A set of n convex hexagons with a bipartite intersection graph and $3n - \Theta(1)$ tangencies. This double grid is "wrapped" around a cylinder such that the pairs of *black and hollow* points coincide, then projected back to the plane

Similarly, we have that $\mathrm{ch}_2(b) \geq \frac{l}{15}$, where $l \in \{1,2\}$ is the number of 5-faces from F_1 that are adjacent to b. It follows that if $\mathrm{ch}_2(t) = -\frac{2}{15}$, then in Round 4 a and b can each send $\frac{1}{15}$ units to t and so $\mathrm{ch}_4(t) = 0$. □

Lemma 2.8 now follows from Propositions 2.11 and 2.12. □

The upper bound in Theorem 3 follows from (3) and Lemma 2.8.

Proposition 2.13. *For any n there is a set $\mathcal{C}$ of n convex regions in the plane, such that $\mathcal{I}(\mathcal{C})$ is bipartite and $t(\mathcal{C}) \geq 3n - \Theta(1)$.*

Proof. We use the same construction of Pach et al. [5] (see Fig. 7), which yields a set $\mathcal{C}$ of n convex hexagons with $t(\mathcal{C}) \geq 3n - \Theta(\sqrt{n})$. However, we observe that one can take this double hexagonal grid with a constant number of rows, ”wrap” it around a cylinder, and then project it back to the plane. Observe that in such a construction, all but a constant number of hexagons touch exactly six other hexagons.

□

References

1. P.K. Agarwal, E. Nevo, J. Pach, R. Pinchasi, M. Sharir, S. Smorodinsky, Lenses in arrangements of pseudo-circles and their applications. J. ACM **51**, 139–186 (2004)
2. N. Alon, H. Last, R. Pinchasi, M. Sharir, On the complexity of arrangements of circles in the plane. Discr. Comput. Geom. **26**, 465–492 (2001)
3. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, Berlin, 2005)
4. P. Erdős, On sets of distances of *n* points. Am. Math. Mon. **53**, 248–250 (1946)
5. J. Pach, A. Suk, M. Treml, Tangencies between families of disjoint regions in the plane, in *Proceedings of the 26th ACM Symposium on Computational Geometry (SoCG 2010)*, Snowbird, June 2010, pp. 423–428
6. J. Spencer, E. Szemerédi, W.T. Trotter, Unit distances in the Euclidean plane, in *Graph Theory and Combinatorics*, ed. by B. Bollobás (Academic, London, 1984), pp. 293–303

Thirty Essays on Geometric Graph Theory*

Greg Aloupis, Brad Ballinger, Sébastien Collette, Stefan Langerman, Attila Pór, and David R. Wood

Abstract This paper studies problems related to visibility among points in the plane. A point x *blocks* two points v and w if x is in the interior of the line segment $\overline{vw}$. A set of points P is k-*blocked* if each point in P is assigned one of k colors, such that distinct points $v, w \in P$ are assigned the same color if and only if some other point in P blocks v and w. The focus of this paper is the conjecture that each k-blocked set has bounded size (as a function of k). Results in the literature imply that every 2-blocked set has at most 3 points, and every 3-blocked set has at most 6 points. We prove that every 4-blocked set has at most 12 points, and that this

*This is the full version of an extended abstract presented at the 26th European Workshop on Computational Geometry (EuroCG'10).

G. Aloupis • S. Collette
Chargé de Recherches du F.R.S.-FNRS, Département d'Informatique,
Université Libre de Bruxelles, Brussels, Belgium
e-mail: aloupis.greg@gmail.com; sebastien.collette@ulb.ac.be

B. Ballinger
Department of Mathematics, Humboldt State University, Arcata, CA, USA
e-mail: bjb86@humboldt.edu

S. Langerman
Maître de Recherches du F.R.S.-FNRS, Département d'Informatique,
Université Libre de Bruxelles, Brussels, Belgium
e-mail: stefan.langerman@ulb.ac.be

A. Pór
Department of Mathematics, Western Kentucky University, Bowling Green, KY, USA
e-mail: attila.por@wku.edu

D.R. Wood (✉)
Department of Mathematics and Statistics, The University of Melbourne, Melbourne, Australia
e-mail: woodd@unimelb.edu.au.

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_4,

bound is tight. In fact, we characterize all sets $\{n_1, n_2, n_3, n_4\}$ such that some 4-blocked set has exactly n_i points in the ith color class. Among other results, for infinitely many values of k, we construct k-blocked sets with $k^{1.79\ldots}$ points.

1 Introduction

This paper studies problems related to visibility and blocking in sets of colored points in the plane. A point x *blocks* two points v and w if x is in the interior of the line segment $\overline{vw}$. Let P be a finite set of points in the plane. Two points v and w are *visible* with respect to P if no point in P blocks v and w. The *visibility graph* of P has vertex set P, where two distinct points $v, w \in P$ are adjacent if and only if they are visible with respect to P. A point set B *blocks* P if $P \cap B = \emptyset$ and for all distinct $v, w \in P$ there is a point in B that blocks v and w. That is, no two points in P are visible with respect to $P \cup B$, or alternatively, P is an independent set in the visibility graph of $P \cup B$.

A set of points P is *k-blocked* if each point in P is assigned one of k colors, such that each pair of points $v, w \in P$ are visible with respect to P if and only if v and w are colored differently. Thus, v and w are assigned the same color if and only if some other point in P blocks v and w. A *k-set* is a multiset of k positive integers. For a k-set $\{n_1, \ldots, n_k\}$, we say P is $\{n_1, \ldots, n_k\}$-blocked if it is k-blocked, and for some labeling of the colors by the integers $[k] := \{1, 2, \ldots, k\}$, the ith color class has exactly n_i points, for each $i \in [k]$. Equivalently, P is $\{n_1, \ldots, n_k\}$-blocked if the visibility graph of P is the complete k-partite graph $K(n_1, \ldots, n_k)$. A k-set $\{n_1, \ldots, n_k\}$ is *representable* if there is an $\{n_1, \ldots, n_k\}$-blocked point set. See Fig. 1 for an example.

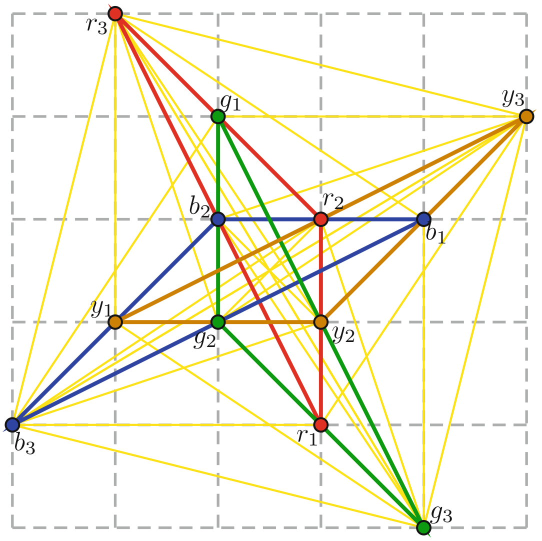


Fig. 1 A $\{3,3,3,3\}$-blocked point set

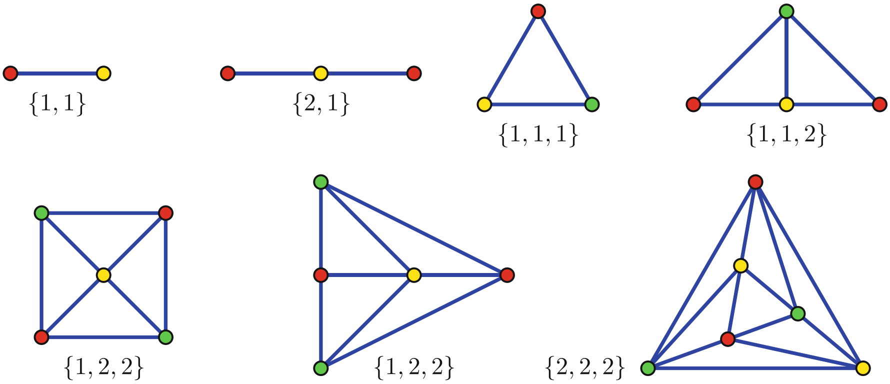


Fig. 2 The 2-blocked and 3-blocked point sets

The following fundamental conjecture regarding k-blocked point sets is the focus of this paper.

Conjecture 1. For each integer k, there is an integer n such that every k-blocked set has at most n points.

As illustrated in Fig. 2, the following theorem is a direct consequence of the characterization of 2- and 3-colorable visibility graphs by Kára et al. [5].

Theorem 2. $\{1,1\}$ *and* $\{1,2\}$ *are the only representable 2-sets, and* $\{1,1,1\}$, $\{1,1,2\}$, $\{1,2,2\}$, *and* $\{2,2,2\}$ *are the only representable 3-sets.*

In particular, every 2-blocked point set has at most three points, and every 3-blocked point set has at most six points. This proves Conjecture 1 for $k \leq 3$.

This paper makes the following contributions. Section 2 introduces some background motivation for the study of k-blocked point sets and observes that results in the literature prove Conjecture 1 for $k = 4$. Section 3 describes methods for constructing k-blocked sets from a given $(k-1)$-blocked set. These methods lead to a characterization of representable k-sets when each color class has at most three points. Section 4 studies the $k = 4$ case in more detail. In particular, we characterize the representable 4-sets, and conclude that the example in Fig. 1 is in fact the largest 4-blocked point set. Section 5 introduces a special class of k-blocked sets (so-called midpoint-blocked sets) that lead to a construction of the largest known k-blocked sets for infinitely many values of k.

Before continuing, we prove some basic properties about k-blocked point sets.

Lemma 3. *At most three points are collinear in every k-blocked point set.*

Proof. Suppose that four points p,q,r,s are collinear in this order. Thus, (p,q,r,s) is an induced path in the visibility graph. Thus, p is not adjacent to r and not adjacent to s. Thus, p, r, and s have the same color. This is a contradiction since r and s are adjacent. Thus, no four points are collinear. □

A set of points P is in *general position* if no three points in P are collinear.

Lemma 4. *Each color class in a k-blocked point set is in general position.*

Proof. Suppose on the contrary that three points from a single color class are collinear. Then no other points are in the same line by Lemma 3. Thus, two of the three points are adjacent, which is a contradiction. □

Note that "convex position" in this paper means "strictly convex position."

2 Some Background Motivation

Much recent research on blockers began with the following conjecture by Kára et al. [5].

Conjecture 5 (Big-Line-Big-Clique Conjecture [5]). For all integers t and ℓ, there is an integer n such that for every finite set P of at least n points in the plane:

- P contains ℓ collinear points, or
- P contains t pairwise visible points (that is, the visibility graph of P contains a t-clique).

Conjecture 5 is true for $t \leq 5$ but is open for $t \geq 6$ or $\ell \geq 4$; see [1,9]. Given that, in general, Conjecture 5 is challenging, Jan Kára suggested the following weakening.

Conjecture 6 ([9]). For all integers t and ℓ, there is an integer n such that for every finite set P of at least n points in the plane:

- P contains ℓ collinear points, or
- The chromatic number of the visibility graph of P is at least t.

Conjecture 5 implies Conjecture 6 since every graph that contains a t-clique has chromatic number at least t.

Proposition 7. *Conjecture 6 with $\ell = 4$ and $t = k+1$ implies Conjecture 1.*

Proof. Assume Conjecture 6 holds for $\ell = 4$ and $t = k+1$. Suppose there is a k-blocked set P of at least n points. By Lemma 3, at most three points are collinear in P. Thus, the first conclusion of Conjecture 6 does not hold. By definition, the visibility graph of P is k-colorable. Thus, the second conclusion of Conjecture 6 does not hold. This contradiction proves that every k-blocked set has fewer than n points, and Conjecture 1 holds. □

Thus, since Conjecture 5 holds for $t \leq 5$, Conjecture 1 holds for $k \leq 4$. In Sect. 4 we take this result much further, by characterizing all representable 4-sets, thus concluding a tight bound on the size of a 4-blocked set.

Part of our interest in blocked point sets comes from the following.

Proposition 8. *For all $k \geq 3$ and $n \geq 2$, the edge set of the k-partite Turán graph $K(n,n,\dots,n)$ can be partitioned into a set of "lines," where*

- *Each line is either an edge or an induced path on three vertices;*
- *Every pair of vertices is in exactly one line; and*
- *For every line L, there is a vertex adjacent to each vertex in L.*

Proof. Let (i,p) be the pth vertex in the ith color class for $i \in \mathbb{Z}_k$ and $p \in \mathbb{Z}_n$ (taken as additive cyclic groups). We introduce three types of lines. First, for $i \in \mathbb{Z}_k$ and distinct $p,q \in \mathbb{Z}_n$, let the triple $\{(i,p),(i+1,p+q),(i,q)\}$ be a line. Second, for $i \in \mathbb{Z}_k$ and $p \in \mathbb{Z}_n$, let the pair $\{(i,p),(i+1,p+p)\}$ be a line. Third, for $p,q \in \mathbb{Z}_n$ and distinct nonconsecutive $i,j \in \mathbb{Z}_k$, let the pair $\{(i,p),(j,q)\}$ be a line. By construction, each line is either an edge or an induced path on three vertices.

Every pair of vertices in the same color class are in exactly one line (of the first type). Consider vertices (i,p) and (j,q) in distinct color classes. First, suppose that i and j are consecutive. Without loss of generality, $j=i+1$. If $q \neq p+p$, then (i,p) and (j,q) are in exactly one line (of the first type). If $q=p+p$, then (i,p) and (j,q) are in exactly one line (of the second type). If i and j are not consecutive, then (i,p) and (j,q) are in exactly one line (of the third type). This proves that every pair of vertices is in exactly one line. Moreover, every edge of $K(n,n,\dots,n)$ is in exactly one line, and the lines partition the edges set.

Since every line L is contained in the union of two color classes, each vertex in neither color class intersecting L is adjacent to each vertex in L. □

Proposition 8 is significant for Conjecture 5 because it says that $K(n,\dots,n)$ behaves like a "visibility space" with no four collinear points, but it has no large clique (for fixed k). Conjecture 1, if true, implies that such a visibility space is not "geometrically representable" for large n.

Let $b(n)$ be the minimum integer such that some set of n points in the plane in general position is blocked by some set of $b(n)$ points. Matoušek [6] proved that $b(n) \geq 2n-3$. Dumitrescu et al. [2] improved this bound to $b(n) \geq (\frac{25}{8}-o(1))n$. Many authors have conjectured or stated as an open problem that $b(n)$ is superlinear.

Conjecture 9 ([2, 6, 8, 9]). $\frac{b(n)}{n} \to \infty$ as $n \to \infty$.

Pór and Wood [9] proved that Conjecture 9 implies Conjecture 6, and thus implies Conjecture 1. That Conjecture 1 is implied by a number of other well-known conjectures, yet remains challenging, adds to its interest.

3 *k*-Blocked Sets with Small Color Classes

We now describe some methods for building blocked point sets from smaller blocked point sets.

Lemma 10. *Let G be a visibility graph. Let $i \in \{1,2,3\}$. Furthermore, suppose that if $i \geq 2$, then $V(G) \neq \emptyset$, and if $i=3$, then not all the vertices of G are collinear. Let G_i be the graph obtained from G by adding an independent set of i new vertices, each adjacent to every vertex in G. Then G_1, G_2, and G_3 are visibility graphs.*

Proof. For distinct points p and q, let $\overleftarrow{pq}$ denote the ray that is (1) contained in the line through p and q, (2) starting at p, and (3) not containing q. Let $\mathcal{L}$ be the union of the set of lines containing at least two vertices in G.

$i = 1$: Since $\mathcal{L}$ is the union of finitely many lines, there is a point $p \notin \mathcal{L}$. Thus, p is visible from every vertex of G. By adding a new vertex at p, we obtain a representation of G_1 as a visibility graph.

$i = 2$: Let p be a point not in $\mathcal{L}$. Let v be a vertex of G. Each line in $\mathcal{L}$ intersects $\overleftarrow{vp}$ in at most one point. Thus, $\overleftarrow{vp} \setminus \mathcal{L} \neq \emptyset$. Let q be a point in $\overleftarrow{vp} \setminus \mathcal{L}$. Thus, p and q are visible from every vertex of G, but p and q are blocked by v. By adding new vertices at p and q, we obtain a representation of G_2 as a visibility graph.

$i = 3$: Let u, v, w be noncollinear vertices in G. Let p be a point not in $\mathcal{L}$ and not in the convex hull of $\{u, v, w\}$. Without loss of generality, $\overline{uv} \cap \overline{pw} \neq \emptyset$. There are infinitely many pairs of points $q \in \overleftarrow{up}$ and $r \in \overleftarrow{vp}$ such that w blocks q and r. Thus, there are such q and r both not in $\mathcal{L}$. By construction, u blocks p and q, and v blocks p and r. By adding new vertices at p, q, and r, we obtain a representation of G_3 as a visibility graph. □

Since no (≥ 3)-blocked set is collinear, Lemma 10 implies

Corollary 11. *If $k \geq 4$ and $\{n_1, \dots, n_{k-1}\}$ is representable and $n_k \in \{1, 2, 3\}$, then $\{n_1, \dots, n_{k-1}, n_k\}$ is representable.*

The representable (≤ 3)-sets were characterized in Theorem 2. In each case, each color class has at most three vertices. Now we characterize the representable (≥ 4)-sets, assuming that each color class has at most three vertices.

Proposition 12. *$\{n_1, \dots, n_k\}$ is representable whenever $k \geq 4$ and each $n_i \leq 3$, except for $\{1, 3, 3, 3\}$.*

Proof. Say the k-set $\{n_1, \dots, n_k\}$ *contains* the $(k-1)$-set $\{n_1, \dots, n_{i-1}, n_{i+1}, \dots, n_k\}$ for each $i \in [k]$. We proceed by induction on $k \geq 4$. For the base case, assume that $k = 4$. If $\{n_1, n_2, n_3, n_4\}$ contains $\{1, 1, 1\}$, $\{1, 1, 2\}$, $\{1, 2, 2\}$, or $\{2, 2, 2\}$, then $\{n_1, n_2, n_3, n_4\}$ is representable by Theorem 2 and Corollary 11. In each remaining case, the following table describes a construction, except for $\{1, 3, 3, 3\}$, which is not representable, as proved in Lemma 18.

$\{1,1,1,x\}$	Contains $\{1,1,1\}$	$\{1,1,2,x\}$	Contains $\{1,1,2\}$
$\{1,1,3,3\}$	Figure 1 minus $\{r_1, g_3, r_3, g_1\}$	$\{1,2,2,x\}$	Contains $\{1,2,2\}$
$\{1,2,3,3\}$	Figure 1 minus $\{g_1, g_3, r_3\}$	$\{2,2,2,x\}$	Contains $\{2,2,2\}$
$\{2,2,3,3\}$	Figure 1 minus $\{g_3, r_3\}$	$\{2,3,3,3\}$	Figure 1 minus g_3
$\{3,3,3,3\}$	Figure 1		

Now assume that $k \geq 5$: Consider a k-tuple $\{n_1, \dots, n_k\}$ with each $n_i \leq 3$. Say $n_1 \leq \cdots \leq n_k$. If $\{n_1, \dots, n_{k-1}\} \neq \{1, 3, 3, 3\}$, then by induction $\{n_1, \dots, n_{k-1}\}$ is representable, which implies that $\{n_1, \dots, n_k\}$ is representable by Corollary 11. Otherwise, $k = 5$ and $\{n_1, \dots, n_5\} = \{1, 3, 3, 3, 3\}$, which contains the representable 4-tuple $\{3, 3, 3, 3\}$, implying $\{n_1, \dots, n_5\}$ is representable by Corollary 11. □

4 4-Blocked Point Sets

As we saw in Sect. 2, Conjecture 1 holds for $k \leq 4$. In this section we study 4-blocked point sets in more detail. First, we derive explicit bounds on the size of 4-blocked sets from other results in the literature. Then, following a more detailed approach, we characterize all representable 4-sets, to conclude a tight bound on the size of 4-blocked sets.

Proposition 13. *Every* 4*-blocked set has less than* 2^{790} *points.*

Proof. Abel et al. [1] proved that every set of at least $\mathrm{ES}(\frac{(2\ell-1)^{\ell}-1}{2\ell-2})$ points in the plane contains ℓ collinear points or an empty convex pentagon, where $\mathrm{ES}(k)$ is the minimum integer such that every set of at least $\mathrm{ES}(k)$ points in general position in the plane contains k points in convex position. Let P be a 4-blocked set. The visibility graph of P is 4-colorable, and thus contains no empty convex pentagon. By Lemma 3, at most three points in P are collinear. Thus, $|P| \leq \mathrm{ES}(400) - 1$ by the above result with $\ell = 4$. Tóth and Valtr [12] proved that $\mathrm{ES}(k) \leq \binom{2k-5}{k-2} + 1$. Hence, $|P| \leq \binom{795}{398} < 2^{790}$. □

Lemma 14. *If P is a blocked set of n points with m points in the largest color class, then* $n \geq 3m - 3$ *and* $n \geq (\frac{33}{8} - o(1))m$.

Proof. If S is the largest color class, then $P - S$ blocks S. By Lemma 4, S is in general position. By the results of Matoušek [6] and Dumitrescu et al. [2] mentioned in Sect. 2, $n - m \geq 2m - 3$ and $n - m \geq (\frac{25}{8} - o(1))m$. □

Proposition 15. *Every* 4*-blocked set has fewer than* 2^{578} *points.*

Proof. Let P be a 4-blocked set of n points. Let S be the largest color class in P. Let $m := |S| \geq \frac{n}{4}$. By Lemma 14, $n \geq (\frac{33}{8} - o(1))m \geq (\frac{33}{32} - o(1))n$. Thus, $o(n) \geq \frac{n}{32}$. Hence, n is bounded. A precise bound is obtained as follows. Dumitrescu et al. [2] proved that S needs at least $\frac{25m}{8} - \frac{25m}{2\ln m} - \frac{25}{8}$ blockers, which come from the other color classes. Thus, $n - m \geq \frac{25m}{8} - \frac{25m}{2\ln m} - \frac{25}{8}$, implying $n \geq \frac{33m}{8} - \frac{25m}{2\ln m} - \frac{25}{8}$. Since $\frac{n}{4} \leq m \leq n$, we have $\frac{25n}{2\ln n} + \frac{25}{8} \geq \frac{n}{32}$. It follows that $n \leq 2^{578}$. □

The next result is the simplest known proof that every 4-blocked point set has bounded size.

Proposition 16. *Every* 4*-blocked set has at most* 36 *points.*

Proof. Let P be a 4-blocked set. Suppose that $|P| \geq 37$. Let S be the largest color class. Thus, $|S| \geq 10$. By Lemma 4, S is in general position. By a theorem of Harborth [4], some 5-point subset $K \subseteq S$ is the vertex-set of an empty convex pentagon $\mathrm{conv}(K)$. Let $T := P \cap (\mathrm{conv}(K) - K)$. Since $\mathrm{conv}(K)$ is empty with respect to S, each point in T is not in S. Thus, T is 3-blocked. K needs at least eight blockers (five blockers for the edges on the boundary of $\mathrm{conv}(K)$, and three blockers for the chords of $\mathrm{conv}(K)$). Thus, $|T| \geq 8$. But every 3-blocked set has at most six points, which is a contradiction. Hence, $|P| \leq 36$. □

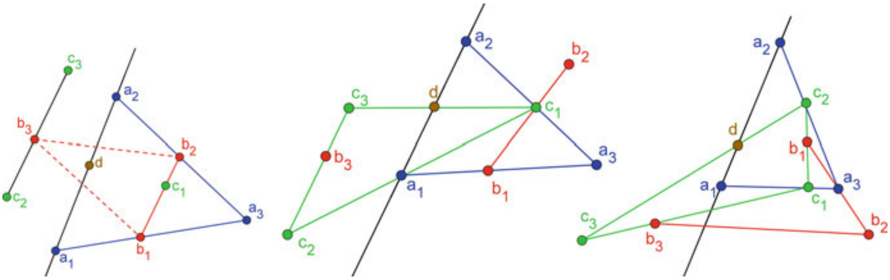


Fig. 3 Cases 1.1, 1.2, and 2 in Lemma 18

We now set out to characterize all representable 4-sets. We need a few technical lemmas.

Lemma 17. *Let A be a set of three monochromatic points in a 4-blocked set P. Then* $P \cap \text{conv}(A)$ *contains a point from each color class.*

Proof. $P \cap \text{conv}(A)$ contains at least three points in A. If $P \cap \text{conv}(A)$ contains no point from one of the three other color classes, then $P \cap \text{conv}(A)$ is a (≤ 3)-blocked set with three points in one color class (A), contradicting Theorem 2. □

Let $a * b * c$ mean that b blocks a and c, and let $a * b * c * d$ mean that $a * b * c$ and $b * c * d$. This allows us to record the order in which points occur on a line. Let $\overleftrightarrow{ab}$ be the line through two points a and b. For three noncollinear points a, b, c, let Δabc be the open triangle with vertices a, b, c.

Lemma 18. $\{3,3,3,1\}$ *is not representable.*

Proof. Suppose that P is a $\{3,3,3,1\}$-blocked set of points, with color classes $A = \{a_1, a_2, a_3\}$, $B = \{b_1, b_2, b_3\}$, $C = \{c_1, c_2, c_3\}$, and $D = \{d\}$.

If d does not block some monochromatic pair, then $P \setminus D$ is a 3-blocked set of nine points, contradicting Theorem 2. Therefore, d blocks some monochromatic pair, which we may call a_1, a_2. Now $\overleftrightarrow{a_1 a_2}$ divides the remaining seven points of P into two sets, P_1 and P_2, where without loss of generality, $4 \leq |P_1| \leq 7$ and $0 \leq |P_2| \leq 3$. If $|P_1| \geq 6$, then $P_1 \cup \{a_1\}$ is a 3-blocked set of more than six points, contradicting Theorem 2. Thus, $|P_1| = 4$ and $|P_2| = 3$, or $|P_1| = 5$ and $|P_2| = 2$. Consider the following cases, as illustrated in Fig. 3.

Case 1. $|P_1| = 4$ and $|P_2| = 3$.

By Theorem 2, P_1 is a $\{1,1,2\}$-blocked set. Thus, without loss of generality, $P_1 = \{a_3, b_1, b_2, c_1\}$ and $P_2 = \{b_3, c_2, c_3\}$, where $c_2 * b_3 * c_3$. Some point in P_1 blocks b_1 and b_2. If $b_1 * a_3 * b_2$, then neither b_1 nor b_2 can block $\overline{a_3 a_1}$ or $\overline{a_3 a_2}$; thus, c_1 blocks a_3 from both a_1 and a_2, a contradiction. Therefore, $b_1 * c_1 * b_2$. Since two of these three points must block $\overline{a_3 a_1}$ and $\overline{a_3 a_2}$, we may assume that $a_1 * b_1 * a_3$, and either $a_2 * b_2 * a_3$ or $a_2 * c_1 * a_3$.

Case 1.1. $a_2 * b_2 * a_3$: Since $c_2 * b_3 * c_3$, the only possible blockers for $\overline{b_1 b_3}$ are on $\overleftrightarrow{a_1 a_2}$. We cannot have $b_3 * a_1 * b_1$, for then $b_3 * a_1 * b_1 * a_3$. We cannot have $b_3 * a_2 * b_1$, for then $\overleftrightarrow{b_1 b_3}$ separates b_2 from d, implying $d \notin \text{conv}(B)$, contradicting Lemma 17. Therefore, $b_1 * d * b_3$. Similarly, $b_2 * d * b_3$, a contradiction.

Case 1.2. $a_2 * c_1 * a_3$: Thus, a_2 does not block $\overline{c_1 c_2}$ or $\overline{c_1 c_3}$. Therefore, a_1 and d block $\overline{c_1 c_2}$ and $\overline{c_1 c_3}$. Without loss of generality, $c_2 * a_1 * c_1$ and $c_3 * d * c_1$. Since $c_2 * b_3 * c_3$, the blocker for $\overline{b_1 b_3}$ is on the same side of $\overleftrightarrow{c_1 c_3}$ as a_1, and on the same side of $\overleftrightarrow{b_1 b_2}$ as a_1. The only such points are a_1, c_2, and b_3. Thus, a_1 or c_2 blocks $\overline{b_1 b_3}$. Now, c_2 does not block $\overline{b_1 b_3}$, as otherwise $b_1 * c_2 * b_3 * c_3$. Similarly, a_1 does not block $\overline{b_1 b_3}$, as otherwise $b_3 * a_1 * b_1 * a_3$. This contradiction concludes this case.

Case 2. $|P_1| = 5$ and $|P_2| = 2$.

We have $a_3 \in P_1$, as otherwise P_1 is 2-blocked, contradicting Theorem 2. Without loss of generality, $P_1 = \{a_3, b_1, b_2, c_1, c_2\}$ and $P_2 = \{b_3, c_3\}$. Note that a_3 must block at least one of $\overline{b_1 b_2}$ and $\overline{c_1 c_2}$, because $P_1 \setminus \{a_3\}$ is too large to be 2-blocked; however, it cannot block both, for then there would be no valid blockers left for a_3. Thus, without loss of generality, $b_1 * a_3 * b_2$ and $c_1 * b_1 * c_2$. Since $a_3 \in \overline{b_1 b_2}$, neither b_1 nor b_2 blocks $\overline{a_3 a_1}$ or $\overline{a_3 a_2}$. Thus, without loss of generality, $a_1 * c_1 * a_3$ and $a_2 * c_2 * a_3$.

By Lemma 17, $P \cap \text{conv}\{c_1, c_2, c_3\}$ contains some member of A, say a_1. We cannot have $a_1 \in \overline{c_1 c_3}$, for then $c_3 * a_1 * c_1 * a_3$. Therefore, a_1 is on the same side of $\overleftrightarrow{c_1 c_3}$ as c_2; consequently, d and a_2 are, as well. It follows that $c_1 * b_3 * c_3$, and so b_3 sees b_2. □

Lemma 19. *Let P be a 4-blocked set. Suppose that some color class S of P contains a subset K, such that $|K| = 4$ and K is the vertex-set of a convex quadrilateral* $\text{conv}(K)$ *that is empty with respect to S. Then P is $\{4,2,2,1\}$-blocked.*

Proof. Let $T := P \cap (\text{conv}(K) - K)$. Since $\text{conv}(K)$ is empty with respect to S, each point in T is not in S. Thus, T is 3-blocked. K needs at least five blockers (four blockers for the edges on the boundary of $\text{conv}(K)$, and at least one blocker for the chords of $\text{conv}(K)$). The only representable 3-sets with at least five points are $\{2,2,1\}$ and $\{2,2,2\}$. At least four points in T are on the boundary of $\text{conv}(T)$. Every $\{2,2,2\}$-blocked set contains three points on the boundary of the convex hull. Thus, T is $\{2,2,1\}$-blocked. Hence, as illustrated in Fig. 4, one point c in T is at the intersection of the two chords of $\text{conv}(K)$, and exactly one point in T is on each edge of the boundary of $\text{conv}(K)$, such that the points on opposite edges of $\text{conv}(K)$ are collinear with c.

We claim that no other point is in P. Suppose otherwise, and let p be a point in P outside $\text{conv}(K)$ at a minimum distance from $\text{conv}(K)$. Let x be a point in $\text{conv}(K)$ receiving the same color as p. Thus, p and x are blocked by some point b in $\text{conv}(K)$. Hence, b and x are collinear with no other point in $P \cap \text{conv}(K)$, so $x \neq c$. Since p sees the nearest vertex in $\text{conv}(K)$, we also have $x \notin K$. Thus, x is in the interior of one of the edges of the boundary of $\text{conv}(K)$. Let y be the point in $\text{conv}(K)$ receiving the same color as x. Thus, x and y are on opposite edges of the boundary of $\text{conv}(K)$. Hence, p and y receive the same color, implying p and y are blocked. Since p is at a

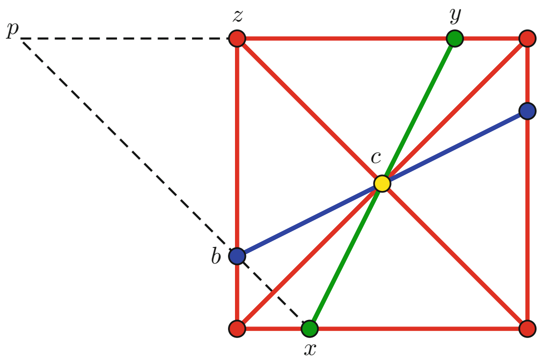


Fig. 4 A $\{4,2,2,1\}$-blocked point set

minimum distance from $\text{conv}(K)$, this blocker must be a point z of $\text{conv}(K)$. Having both $p*b*x$ and $p*z*y$ implies that p,z,y and one other point of $\text{conv}(K)$ are four collinear points, which contradicts Lemma 3. Hence, no other point is in P, and P is $\{4,2,2,1\}$-blocked. □

Lemma 19 has the following corollary (let $K := S$).

Corollary 20. *Let P be a 4-blocked set. Suppose that some color class S consists of exactly four points in convex position. Then P is $\{4,2,2,1\}$-blocked.*

The next lemma is a key step in our characterization of representable 4-sets.

Lemma 21. *Each color class in a 4-blocked point set has at most four points.*

Proof. Suppose that some 4-blocked point set P has a color class S with at least five points. Esther Klein [3] proved that every set of at least five points in general position in the plane contains an empty quadrilateral. By Lemma 4, S is in general position. Thus, S contains a subset K, such that $|K| = 4$ and K is the vertex set of a convex quadrilateral $\text{conv}(K)$ that is empty with respect to S. By Lemma 19, P is $\{4,2,2,1\}$-blocked, which is the desired contradiction. □

Lemma 22. *Let P be a 4-blocked point set with color classes A,B,C,D. Suppose that no color class consists of exactly four points in convex position (that is, Corollary 20 is not applicable). Furthermore, suppose that some color class A consists of exactly four points in nonconvex position. Then P is $\{4,2,2,2\}$-blocked, as in Fig. 5 for example.*

Proof. By Lemma 21, each color class has at most four points. By assumption, every 4-point color class is in nonconvex position. We may assume that A is minimal in the sense that no other 4-point color class is within $\text{conv}(A)$.

Let $Q := P \cap \text{conv}(A)$. Thus, Q is 4-blocked, and one color class is A. By the minimality of A, each other color class in Q has at most three points. We first prove that Q is $\{4,2,2,2\}$-blocked, and then show that this implies that $P = Q$.

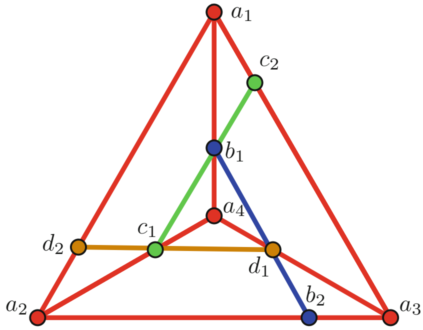


Fig. 5 A $\{4,2,2,2\}$-blocked point set

Let $A = \{a_1, a_2, a_3, a_4\}$, where a_4 is the interior point of $\text{conv}(A)$. Note that the edges with endpoints in A divide $\text{conv}(A)$ into three triangles with disjoint interiors. By Lemma 17, each color class of Q is represented in each of these triangles; this requires at least two points of each color (one of which could sit on the edge shared by two triangles).

We name a point with reference to its color class, such as $b_1 \in B$; or we name a point with reference to its position.

Let $Q' := Q \setminus A$. Let x_{ij} be the unique member of Q' that blocks $\overline{a_i a_j}$, for $1 \leq i < j \leq 4$. This accounts for exactly six points of Q'. For each case below we prove that Q' contains four mutually visible points (with respect to Q), which is a contradiction since Q' is 3-colored.

Q is not $\{4,3,3,3\}$-blocked, as otherwise we could delete the three outer members of A to represent $\{3,3,3,1\}$, which contradicts Lemma 18. Thus, Q is $\{4,3,2,2\}$-blocked, $\{4,3,3,2\}$-blocked, or $\{4,2,2,2\}$-blocked.

First, suppose that Q is $\{4,3,2,2\}$-blocked. Then Q' consists of six points x_{ij} and one additional point, y. The following three cases arise, as illustrated in Fig. 6.

Case 1. y blocks two points x_{i4} and x_{j4}: Without loss of generality, $x_{24} * y * x_{34}$. Now y is on one side or the other of $\overleftrightarrow{x_{14}a_4}$, and therefore y sees x_{12} or x_{13}. Without loss of generality, y sees x_{12}. Thus, x_{12}, x_{14}, x_{24}, and y are mutually visible points in Q'.

Case 2. y is collinear with, but does not block, two points x_{i4} and x_{j4}. Without loss of generality, $y * x_{24} * x_{34}$. Now x_{12} is on one side or the other of $\overleftrightarrow{yx_{24}}$; thus, x_{12} sees some $p \in \{x_{14}, x_{23}\}$, which is mutually visible with x_{12}, x_{24} and y; thus, we have four mutually visible points in Q'.

Case 3. y is not collinear with two points x_{i4} and x_{j4}: Then y sees all such points; thus, x_{14}, x_{24}, x_{34}, and y are mutually visible in Q'.

Now suppose that Q is $\{4,3,3,2\}$-blocked. Then Q' consists of the six points x_{ij} and two additional points, y_1 and y_2. Let $\delta = \{x_{14}, x_{24}, x_{34}\}$ and let δ_2 be the set of segments with both endpoints in δ. How many of the points y_i block members of δ_2? The following cases arise.

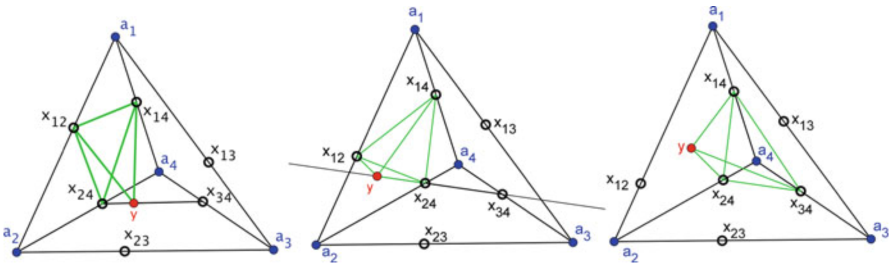


Fig. 6 Cases 1–3 when Q is $\{4,3,2,2\}$-blocked

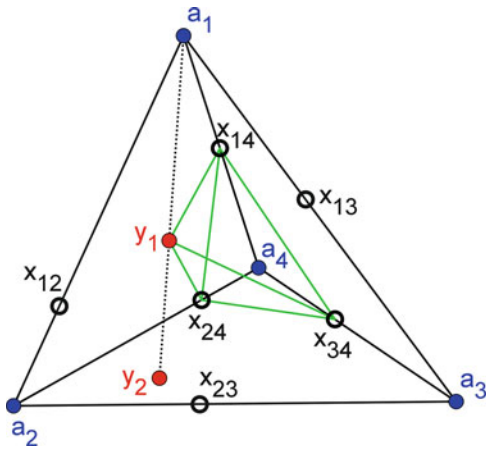


Fig. 7 Case 1 when Q is $\{4,3,3,2\}$-blocked

Case 1. Neither y_1 nor y_2 blocks a segment in δ_2 (see Fig. 7).

Since y_1 and y_2 cannot block each other, without loss of generality, y_2 does not block y_1. Thus, δ and y_1 give four mutually visible points in Q'.

Case 2. At least one of y_1 and y_2 blocks a segment in δ_2.

Without loss of generality, y_1 blocks a segment in δ_2 and $x_{24} * y_1 * x_{34}$. Neither a_1 nor a_4 belongs to $\overleftrightarrow{x_{14}y_1}$, because otherwise *both* do (since $a_1 * x_{14} * a_4$), yielding four collinear points. Since $y_1 \in \Delta a_2a_3a_4 \subset \Delta a_2a_3x_{14}$, we know that a_2 and a_3 are on opposite sides of $\overleftrightarrow{x_{14}y_1}$ by the Crossbar Theorem. Thus, without loss of generality, a_4 is on the same side of $\overleftrightarrow{x_{14}y_1}$ as x_{24}—call this side "west" and the other side "east."

We now show that $\{x_{13}, x_{14}, x_{34}, y_1\}$ is a set of four points in general position. If not, then three of these points are collinear. Observe that x_{13}, x_{14}, and x_{34} are noncollinear. Thus, y_1 is one of the three collinear points. Since $\overleftrightarrow{x_{34}y_1}$ already contains a third point, namely x_{24}, it contains neither x_{13} nor x_{14} by Lemma 3. Thus, our three collinear points are $\{x_{13}, x_{14}, y_1\}$. Since $x_{24} * y_1 * x_{34}$, the point x_{34} is east of $\overleftrightarrow{x_{14}y_1}$. Thus, a_1 and a_3 are also east of $\overleftrightarrow{x_{14}y_1}$, so x_{13} must be as well. Therefore, x_{13} is not collinear with x_{14} and y_1. Thus, $\{x_{13}, x_{14}, x_{34}, y_1\}$ is a set of four points in general position.

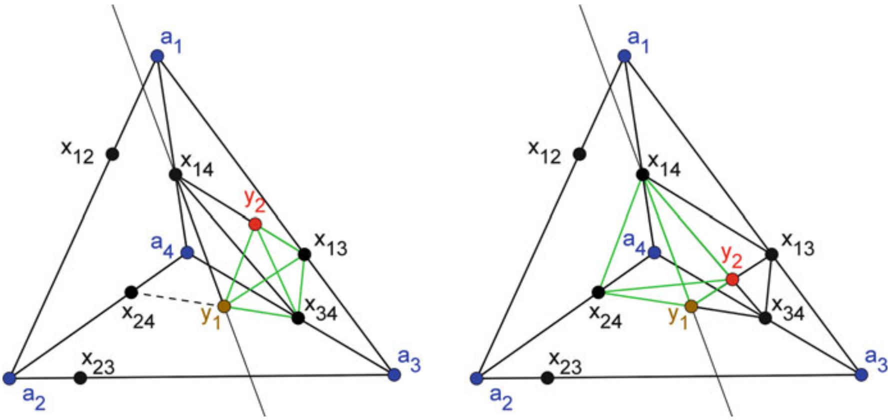


Fig. 8 Case 2a when Q is $\{4,3,3,2\}$-blocked

Since Q' does not contain four mutually visible points, there is some blocker in the midst of $\{x_{13},x_{14},x_{34},y_1\}$. Because a_4 and x_{24} are west of $\overleftrightarrow{x_{14}y_1}$, the only possible blocker—which must be an interior point of conv(Q)—is y_2, which must therefore belong to conv$(x_{13},x_{14},x_{34},y_1)$.

Case 2a. $\{x_{13},x_{14},x_{34},y_1\}$ is in convex position (see Fig. 8).

If y_1 and x_{13} are on the same side of $\overleftrightarrow{x_{14}x_{34}}$, then y_1 and x_{24} would be on opposite sides of $\overleftrightarrow{x_{14}x_{34}}$, implying y_1 would not block $\overline{x_{24}x_{34}}$, which is a contradiction. Thus, y_1 and x_{13} are on opposite sides of $\overleftrightarrow{x_{14}x_{34}}$. It follows that conv$(x_{13},x_{14},x_{34},y_1)$ has chords $\overline{x_{14}x_{34}}$ and $\overline{x_{13}y_1}$. If y_2 blocks just one edge between $\{x_{13},x_{14},x_{34},y_1\}$ (as in Fig. 8, left), then we may substitute y_2 for one endpoint of that edge to build a set of four mutually visible points in Q'. Thus, y_2 must block two edges between $\{x_{13},x_{14},x_{34},y_1\}$ (as in Fig. 8, right). Hence, $x_{14} * y_2 * x_{34}$ and $y_1 * y_2 * x_{13}$. Now (1) a_4 is blocked by both x_{14} and x_{24}, and therefore blocks neither of them; (2) a_4 cannot block $\overline{y_1y_2}$, which is on the other side of $\overleftrightarrow{x_{14}y_1}$; (3) x_{34} is blocked by both y_1 and y_2, and therefore cannot block either of them; (4) x_{34} cannot block $\overline{x_{14}x_{24}}$, which is on the other side of $\overleftrightarrow{x_{14}y_1}$. Thus, $\{x_{14},x_{24},y_1,y_2\}$ is a set of four mutually visible points in Q'.

Case 2b. $\{x_{13},x_{14},x_{34},y_1\}$ is not in convex position (see Fig. 9).

One of these four points lies inside the triangle formed by the other three. This interior point cannot be x_{13}, since it is on the boundary of conv(A). Nor can it be x_{14} or y_1, as we previously showed that x_{13} and x_{34} are on the same side (east) of $\overleftrightarrow{x_{14}y_1}$. Thus, x_{34} is in $\Delta x_{13}x_{14}y_1$. Hence, x_{13} and x_{14} are on opposite sides of $\overleftrightarrow{x_{34}y_1}$. That is, x_{13} and x_{23} are on the same side of $\overleftrightarrow{x_{34}y_1}$. This implies that x_{24} and y_1 are boundary points of conv$(x_{23},x_{24},y_1,x_{13})$. Moreover, x_{13} and x_{23} must also be boundary points of conv$(x_{23},x_{24},y_1,x_{13})$, as they are boundary points of conv(A).

Now $\{x_{23},x_{24},y_1,x_{13}\}$ is a set of four points in convex position. Note that conv$(x_{13},x_{14},x_{34},y_1)$ meets conv$(x_{23},x_{24},y_1,x_{13})$ along only a single edge, $\overline{y_1x_{13}}$.

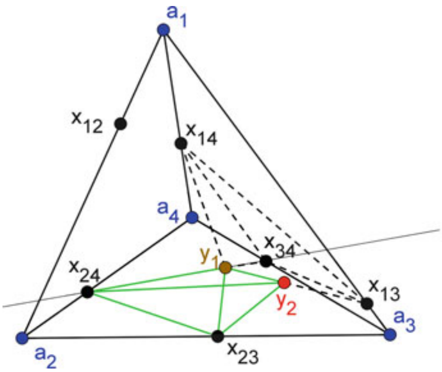


Fig. 9 Case 2b when Q is $\{4,3,3,2\}$-blocked

If y_2 blocks $\overline{y_1x_{13}}$, then $\{x_{23},x_{24},y_1,y_2\}$ is a set of four mutually visible points in Q' (as in Fig. 9, left). If y_2 does not block $\overline{y_1x_{13}}$, then $\{x_{23},x_{24},y_1,x_{13}\}$ is a set of four mutually visible points in Q' (as in Fig. 9, right).

The only remaining case is that Q is $\{4,2,2,2\}$-blocked. This is possible, as illustrated in Fig. 5. We now show that this point set is essentially the only $\{4,2,2,2\}$-blocked set, up to betweenness-preserving deformations. There are exactly enough points in color classes B, C, and D to block all edges between points in A. As each of the three A-triangles with point a_4 must contain a representative from each of the other three color classes, it follows that of the six A-edges, B blocks one interior edge and the opposite boundary edge, and likewise for C and D. Without loss of generality $b_1 = x_{14}$, $b_2 = x_{23}$, $c_1 = x_{24}$, $c_2 = x_{13}$, $d_1 = x_{34}$, and $d_2 = x_{12}$. Since b_1 blocks a_4, a_4 cannot block b_1. Since $\overline{b_1b_2}$ can be blocked only by an interior point of conv(A), it follows that either c_1 or d_1 blocks $\overline{b_1b_2}$. As these cases are symmetric, we may choose $c_1 \in \overline{b_1b_2}$. Now b_1 cannot block c_1, so c_1 must be blocked by d_1, which must in turn be blocked by b_1.

Now we show that $P = Q$. (This basically says that the point set in Fig. 5 cannot be extended without introducing a new color.) Suppose to the contrary that $P \setminus \text{conv}(A) \neq \emptyset$. Let x be a point in $P \setminus \text{conv}(A)$ closest to conv(A). Thus, x sees a vertex of conv(A), and $x \notin A$. Without loss of generality, $x \in B$. Recall that b_2 is on the line $\overleftrightarrow{a_2a_3}$. Thus, x is in the same half-plane determined by $\overleftrightarrow{a_2a_3}$ as the rest of Q, as otherwise b_2 would see $x \in B$. Which point blocks $\overline{b_1x}$? Not a_2 or a_3, for this would put x in the wrong half-plane. Not a_1, a_4, c_1, or c_2, since b_1 blocks each of these. Not d_1, which is on $\overline{b_1b_2}$. Therefore, $\overline{b_1x}$ can be blocked only by d_2. Since $d_1 \in \overline{b_1b_2}$ and $d_2 \in \overline{b_1x}$, it follows that b_2 and x (and any blocker between them) are on the same side of $\overleftrightarrow{d_1d_2}$. But the only other points of Q in that open half-plane are a_2 and a_3, which cannot block b_2. Thus, x sees b_2, which is a contradiction. Thus, $P = Q$ and P is $\{4,2,2,2\}$-blocked. □

We now prove the main theorem of this section.

Theorem 23. *A 4-set* $\{a,b,c,d\}$ *is representable if and only if*

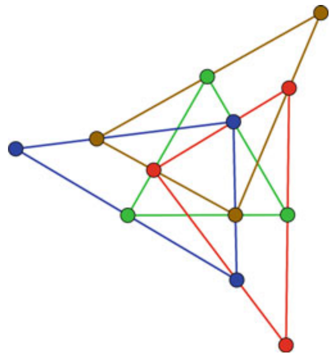


Fig. 10 Another $\{3,3,3,3\}$-blocked point set

- $\{a,b,c,d\} = \{4,2,2,1\}$, *or*
- $\{a,b,c,d\} = \{4,2,2,2\}$, *or*
- *All of* $a,b,c,d \leq 3$ *except for* $\{3,3,3,1\}$.

Proof. Figures 4 and 5 respectively show $\{4,2,2,1\}$-blocked and $\{4,2,2,2\}$-blocked point sets. When $a,b,c,d \leq 3$, the required constructions are described in Proposition 12. Now we prove that these are the only representable 4-sets. Let P be a 4-blocked point set. By Lemma 21, each color class has at most four points. Let S be the largest color class. If $|S| \leq 3$, then we are done by Proposition 12. Now assume that $|S| = 4$. If S is in nonconvex position, then P is $\{4,2,2,2\}$-blocked by Lemma 22. If S is in convex position, then P is $\{4,2,2,1\}$-blocked by Corollary 20. □

Corollary 24. *Every 4-blocked set has at most 12 points, and there is a 4-blocked set with 12 points.*

Note that in addition to the $\{3,3,3,3\}$-blocked set shown in Fig. 1, there is a different $\{3,3,3,3\}$-blocked point set, as illustrated in Fig. 10.

5 Midpoint-Blocked Point Sets

A k-blocked point set P is *k-midpoint-blocked* if for each monochromatic pair of distinct points $v,w \in P$ the midpoint of $\overline{vw}$ is in P. Of course, the midpoint of $\overline{vw}$ blocks v and w. A point set P is *$\{n_1,\dots,n_k\}$-midpoint-blocked* if it is $\{n_1,\dots,n_k\}$-blocked and k-midpoint-blocked. For example, the point set in Fig. 1 is $\{3,3,3,3\}$-midpoint-blocked.

Another interesting example is the projection[1] of $[3]^d$. With $d = 1$ this point set is $\{2,1\}$-blocked, with $d = 2$ it is $\{4,2,2,1\}$-blocked, and with $d = 3$ it is

[1]If G is the visibility graph of some point set $P \subseteq \mathbb{R}^d$, then G is the visibility graph of some projection of P to $\mathbb{R}^2$ (since a random projection of P to $\mathbb{R}^2$ is occlusion-free with probability 1).

$\{8,4,4,4,2,2,2,1\}$-blocked. In general, each set of points with exactly the same set of coordinates equal to 2 is a color class. Each color class of points with exactly i coordinates equal to 2 has size 2^{d-i}, and there are $\binom{d}{i}$ such color classes. Hence, $[3]^d$ is $\{\binom{d}{i} \times 2^{d-i} : i \in [0,d]\}$-midpoint-blocked and 2^d-midpoint-blocked.

We now prove Conjecture 1 when restricted to k-midpoint-blocked point sets. (Finally, we have weakened Conjecture 5 to something provable!)

A. Hernández-Barrera, F. Hurtado, J. Urrutia, and C. Zamora (On the midpoints of a plane point set, 2001, unpublished manuscript) introduced the following definition. Let $m(n)$ be the minimum number of midpoints determined by some set of n points in general position in the plane. Since the midpoint of $\overline{vw}$ blocks v and w, we have $b(n) \le m(n)$. A. Hernández-Barrera, F. Hurtado, J. Urrutia, and C. Zamora (On the midpoints of a plane point set, 2001, unpublished manuscript) constructed a set of n points in general position in the plane that determine at most $cn^{\log 3}$ midpoints for some constant c. (All logarithms here are binary.) Thus, $b(n) \le m(n) \le cn^{\log 3} = cn^{1.585\cdots}$. This upper bound was improved independently by Pach [7] and Stanchescu [11] (and later by Matoušek [6]) to

$$b(n) \le m(n) \le nc^{\sqrt{\log n}}.$$

A. Hernández-Barrera, F. Hurtado, J. Urrutia, and C. Zamora (On the midpoints of a plane point set, 2001, unpublished manuscript) conjectured that $m(n)$ is superlinear (that is, $\frac{m(n)}{n} \to \infty$ as $n \to \infty$), which was verified by Pach [7]. More precise bounds were obtained by Stanchescu [11] and Sanders [10]. Applying the latest results on Freiman's Theorem, Pór and Wood [9] observed that for all $\varepsilon > 0$ there is an integer $N(\varepsilon)$ such that $m(n) \ge n(\log n)^{4/11-\varepsilon}$ for all $n \ge N(\varepsilon)$.

Theorem 25. *For each k, there is an integer n such that every k-midpoint-blocked set has at most n points. More precisely, for all $\varepsilon > 0$ and for all k, every k-midpoint-blocked set has at most $k \max\{N(\varepsilon), 2^{(k-1)^{11/(4-11\varepsilon)}}\}$ points.*

Proof. Let P be a k-midpoint-blocked set of n points. If $n \le kN(\varepsilon)$, then we are done. Now assume that $\frac{n}{k} > N(\varepsilon)$. Let S be a set of exactly $s := \lceil \frac{n}{k} \rceil$ monochromatic points in P. Thus, S is in general position by Lemma 4. And for every pair of distinct points $v,w \in S$, the midpoint of $\overline{vw}$ is in $P - S$. Thus,

$$\tfrac{n}{k}(\log \tfrac{n}{k})^{4/11-\varepsilon} \le m(s) \le n - s \le n(1 - \tfrac{1}{k}).$$

Hence, $(\log \frac{n}{k})^{4/11-\varepsilon} \le k-1$, implying $n \le k2^{(k-1)^{11/(4-11\varepsilon)}}$. The result follows. □

We now construct k-midpoint-blocked point sets with a "large" number of points. The method is based on the following product of point sets P and Q. For each point $v \in P \cup Q$, let (x_v, y_v) be the coordinates of v. Let $P \times Q$ be the point set $\{(v,w) : v \in P, w \in Q\}$, where (v,w) is at (x_v, y_v, x_w, y_w) in 4-dimensional space. For brevity we do not distinguish between a point in $\mathbb{R}^4$ and its image in an occlusion-free projection of the visibility graph of $P \times Q$ into $\mathbb{R}^2$.

Lemma 26. *If P is a $\{n_1,\dots,n_k\}$-midpoint-blocked point set and Q is a $\{m_1,\dots,m_\ell\}$-midpoint-blocked point set, then $P\times Q$ is $\{n_im_j : i\in[k], j\in[\ell]\}$-midpoint-blocked.*

Proof. Color each point (v,w) in $P\times Q$ by the pair $(\text{col}(v),\text{col}(w))$. Thus, there are n_im_j points for the (i,j)th color class. Consider distinct points (v,w) and (a,b) in $P\times Q$.

Suppose that $\text{col}(v,w)=\text{col}(a,b)$. Thus, $\text{col}(v)=\text{col}(a)$ and $\text{col}(w)=\text{col}(b)$. Since P and Q are midpoint-blocked, $\frac{1}{2}(v+a)\in P$ and $\frac{1}{2}(w+b)\in Q$. Thus, $(\frac{1}{2}(v+a),\frac{1}{2}(w+b))$, which is positioned at $(\frac{1}{2}(x_v+x_a),\frac{1}{2}(y_v+y_a),\frac{1}{2}(x_w+x_b),\frac{1}{2}(y_w+y_b))$, is in $P\times Q$. This point is the midpoint of $\overline{(v,w)(a,b)}$. Thus, (v,w) and (a,b) are blocked by their midpoint in $P\times Q$.

Conversely, suppose that some point $(r,s)\in P\times Q$ blocks (v,w) and (a,b). Thus, $x_r=\alpha x_v+(1-\alpha)x_a$ for some $\alpha\in(0,1)$, $y_r=\beta y_v+(1-\beta)y_a$ for some $\beta\in(0,1)$, $x_s=\delta x_w+(1-\delta)x_b$ for some $\delta\in(0,1)$, and $y_s=\gamma y_w+(1-\gamma)y_b$ for some $\gamma\in(0,1)$. Hence, r blocks v and a in P, and s blocks w and b in Q. Thus, $\text{col}(v)=\text{col}(a)\neq\text{col}(r)$ in P, and $\text{col}(w)=\text{col}(b)\neq\text{col}(s)$ in Q, implying $(\text{col}(v),\text{col}(w))=(\text{col}(a),\text{col}(b))$.

We have shown that two points in $P\times Q$ are blocked if and only if they have the same color. Thus, $P\times Q$ is blocked. Since every blocker is a midpoint, $P\times Q$ is midpoint-blocked. □

Say P is a k-midpoint blocked set of n points. By Lemma 26, the i-fold product $P^i := P\times\cdots\times P$ is a k^i-blocked set of $n^i=(k^i)^{\log_k n}$ points. Taking P to be the $\{3,3,3,3\}$-midpoint-blocked point set in Fig. 1, we obtain the following result:

Theorem 27. *For all k a power of 4, there is a k-blocked set of $k^{\log_4 12}=k^{1.79\dots}$ points.*

This result describes the largest known construction of k-blocked or k-midpoint-blocked point sets. To promote further research, we make the following strong conjectures:

Conjecture 28. Every k-blocked point set has $O(k^2)$ points.

Conjecture 29. In every k-blocked point set there are at most k points in each color class.

Conjecture 29 would be tight for the projection of $[3]^d$ with $k=2^d$. Of course, Conjecture 29 implies Conjecture 28, which implies Conjecture 1.

Acknowledgements This research was initiated at The 24th Bellairs Winter Workshop on Computational Geometry, held in February 2009 at the Bellairs Research Institute of McGill University in Barbados. The authors are grateful to Godfried Toussaint and Erik Demaine for organizing the workshop, and to the other participants for providing a stimulating working environment. Thanks to the reviewer for many helpful suggestions. David R. Wood was supported by a QEII Research Fellowship from the Australian Research Council.

References

1. Z. Abel, B. Ballinger, P. Bose, S. Collette, V. Dujmović, F. Hurtado, S.D. Kominers, S. Langerman, A. Pór, D.R. Wood, Every large point set contains many collinear points or an empty pentagon. Graphs Comb. **27**(1), 47–60 (2011). http://dx.doi.org/10.1007/s00373-010-0957-2. doi:10.1007/s00373-010-0957-2
2. A. Dumitrescu, J. Pach, G. Tóth, A note on blocking visibility between points. Geombinatorics **19**(1), 67–73 (2009). http://www.cs.uwm.edu/faculty/ad/blocking.pdf
3. P. Erdős, G. Szekeres, A combinatorial problem in geometry. Compos. Math. **2**, 464–470 (1935). http://www.numdam.org/item?id=CM_1935__2__463_0
4. H. Harborth, Konvexe Fünfecke in ebenen Punktmengen. Elem. Math. **33**(5), 116–118 (1978)
5. J. Kára, A. Pór, D.R. Wood, On the chromatic number of the visibility graph of a set of points in the plane. Discr. Comput. Geom. **34**(3), 497–506 (2005). http://dx.doi.org/10.1007/s00454-005-1177-z. doi:10.1007/s00454-005-1177-z
6. J. Matoušek, Blocking visibility for points in general position. Discr. Comput. Geom. **42**(2), 219–223 (2009). http://dx.doi.org/10.1007/s00454-009-9185-z. doi:10.1007/s00454-009-9185-z
7. J. Pach, Midpoints of segments induced by a point set. Geombinatorics **13**(2), 98–105 (2003). http://www.math.nyu.edu/~pach/publications/midpoint.ps
8. R. Pinchasi, On some unrelated problems about planar arrangements of lines, in *Workshop II: Combinatorial Geometry. Combinatorics: Methods and Applications in Mathematics and Computer Science*, Institute for Pure and Applied Mathematics, UCLA, 2009. http://11011110.livejournal.com/184816.html.
9. A. Pór, D.R. Wood, On visibility and blockers. J. Comput. Geom. **1**(1), 29–40 (2010) http://www.jocg.org/index.php/jocg/article/view/24
10. T. Sanders, Three-term arithmetic progressions and sumsets. Proc. Edinb. Math. Soc. (2) **52**(1), 211–233 (2009). http://dx.doi.org/10.1017/S0013091506001398. doi:10.1017/S0013091506001398.
11. Y.V. Stanchescu, Planar sets containing no three collinear points and non-averaging sets of integers. Discr. Math. **256**(1–2), 387–395 (2002) http://dx.doi.org/10.1016/S0012-365X(01)00441-1. doi:10.1016/S0012-365X(01)00441-1.
12. G. Tóth, P. Valtr, The Erdős–Szekeres theorem: upper bounds and related results, in *Combinatorial and Computational Geometry*. Math. Sci. Res. Inst. Publ., vol. 52 (Cambridge University Press, Cambridge, 2005), pp. 557–568. http://library.msri.org/books/Book52/files/30toth.pdf

Constrained Tri-Connected Planar Straight Line Graphs

Marwan Al-Jubeh, Gill Barequet, Mashhood Ishaque, Diane L. Souvaine, Csaba D. Tóth, and Andrew Winslow

Abstract It is known that for any set V of $n \geq 4$ points in the plane, not in convex position, there is a 3-connected planar straight line graph $G = (V, E)$ with at most $2n-2$ edges, and this bound is the best possible. We show that the upper bound $|E| \leq 2n$ continues to hold if $G = (V, E)$ is constrained to contain a given graph $G_0 = (V, E_0)$, which is either a 1-factor (i.e., disjoint line segments) or a 2-factor (i.e., a collection of simple polygons), but no edge in E_0 is a proper diagonal of the convex hull of V. Since there are 1- and 2-factors with n vertices for which any 3-connected augmentation has at least $2n-2$ edges, our bound is nearly tight in these cases. We also examine possible conditions under which this bound may be improved, such as when G_0 is a collection of interior-disjoint convex polygons in a triangular container.

1 Introduction

A graph is *k-connected* if it remains connected upon deleting any $k-1$ vertices along with all incident edges. Connectivity augmentation problems are an important area in optimization and network design. The *k-connectivity augmentation* problem asks

M. Al-Jubeh • M. Ishaque • D.L. Souvaine • A. Winslow
Department of Computer Science, Tufts University, Medford, MA, USA
e-mail: maljub01@cs.tufts.edu; mishaque@cs.tufts.edu; dls@cs.tufts.edu; awinsl02@cs.tufts.edu

G. Barequet
Department of Computer Science, Technion, Haifa, Israel

Department of Computer Science, Tufts University, Medford, MA, USA
e-mail: barequet@cs.tufts.edu; barequet@cs.technion.ac.il

C.D. Tóth (✉)
Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada

Department of Computer Science, Tufts University, Medford, MA, USA
e-mail: cdtoth@cs.tufts.edu; cdtoth@ucalgary.ca

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_5,

for the minimum number of edges needed to augment an input graph $G_0 = (V, E_0)$ to a k-connected graph $G = (V, E)$, $E_0 \subseteq E$. In abstract graphs, the connectivity augmentation problem can be solved in $O(|V| + |E|)$ time for $k = 2$ [3, 6, 7, 11], and in polynomial time for any fixed k [9].

Researchers have considered the connectivity augmentation problems over planar graphs where both the input G_0 and the output G have to be planar (that is, they have no K_5 or $K_{3,3}$ minors). Kant and Bodlaender [10] proved that already the 2-connectivity augmentation over planar graphs is NP-hard, and they devised a 2-approximation algorithm that runs in $O(n \log n)$ time. We consider 3-connectivity augmentation over planar *geometric* graphs, where the given straight-line embedding of the input graph has to be preserved.

A planar straight-line graph (for short, PSLG) is a graph with a straight-line embedding in the plane. That is, the vertices are distinct points in the plane and the edges are straight-line segments between the incident endpoints (that do not pass through any other vertices). The *k-connectivity augmentation for PSLGs* asks for the minimum number of edges needed to augment an input PSLG $G_0 = (V, E_0)$ to a k-connected PSLG $G = (V, E)$, $E_0 \subseteq E$. Rutter and Wolff [12] showed that this problem is NP-hard for every integer k, $2 \leq k \leq 5$. Note that the problem is infeasible for $k \geq 6$, since every planar graph has a vertex of degree at most 5. There are two possible approaches to get around the NP-hardness of the augmentation problem: (a) approximation algorithms, as was done for planarity-preserving 2-connectivity augmentation; and (b) proving extremal bounds for the minimum number of edges sufficient for the augmentation in terms of the number of vertices, which we do here.

It is easy to see that for every $n \geq 4$, there is a 3-connected planar graph with n vertices and $\lceil 3n/2 \rceil$ edges, where all but at most one vertex have degree 3. On a set V of $n \geq 4$ points in the plane, however, a 3-connected PSLG may require many more edges. García et al. [4] proved that if $3 \leq h < n$ points lie on the convex hull of V, then it admits a 3-connected PSLG $G = (V, E)$ with at most $\max(\lceil 3n/2 \rceil, n + h - 1) \leq 2n - 2$ edges, and this bound is the best possible. If the points in V are in convex position (that is, $h = n$), then V does not admit any 3-connected PSLG.

Tóth and Valtr [13] characterized the *3-augmentable* planar straight-line graphs, that is, graphs that can be augmented to 3-connected PSLGs. Specifically, a PSLG $G_0 = (V, E_0)$ is 3-augmentable if and only if E_0 does not contain any edge that is a proper diagonal of the convex hull of V. Every 3-augmentable PSLG on n vertices can be augmented to a 3-connected triangulation, which has up to $3n - 6$ edges, but in some cases significantly fewer edges are sufficient. As mentioned above, the 3-connectivity augmentation problem for PSLGs is NP-hard, and no approximation is known. It is also not known how many new edges are sufficient for augmenting *any* 3-augmentable PSLG with n vertices. Such a worst-case bound is known only for edge-connectivity: Al-Jubeh et al. [2] proved recently that every 3-edge-augmentable PSLG with n vertices can be augmented to a 3-edge-connected PSLG by adding at most $2n - 2$ new edges, and this bound is the best possible.

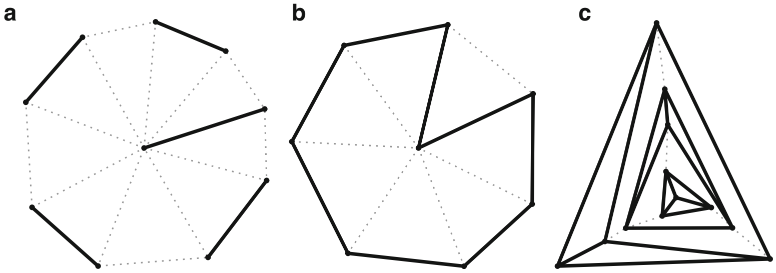


Fig. 1 (**a**, **b**) 1- and 2-regular PSLGs whose only 3-connected augmentation is the wheel graph. (**c**) Nested copies of K_4, for which every 3-connected augmentation has at least $\frac{9}{4}n-3$ edges

Our Results. In the 3-connectivity augmentation problem for PSLGs, we are given a PSLG $G_0 = (V, E_0)$, and asked to augment it to a 3-connected PSLG $G = (V, E)$, $E_0 \subseteq E$. Intuitively, the edges in E_0 are either "useful" for constructing a 3-connected graph or they are "obstacles" that prevent the addition of new edges that would cross them. In this chapter, we explore some classes of 3-augmentable PSLGs with $n \geq 4$ vertices that can be augmented to 3-connected PSLGs that have at most $2n$ edges. Recall that $2n-2$ edges may be necessary even for a completely "unobstructed" input $G_0 = (V, \emptyset)$. We prove that if G_0 is 1-regular (that is, a crossing-free perfect matching) or 2-regular (a collection of pairwise noncrossing simple polygons), then it can be augmented to a 3-connected PSLG that has at most $2n-2$ or $2n$ edges, respectively.

Theorem 1. *Every 1-regular 3-augmentable PSLG $G_0 = (V, E_0)$ with $n \geq 4$ vertices can be augmented to a 3-connected PSLG $G = (V, E)$, $E_0 \subseteq E$, with $|E| \leq 2n-2$ edges.*

Theorem 2. *Every 2-regular 3-augmentable PSLG $G_0 = (V, E)$ with $n \geq 4$ vertices can be augmented to a 3-connected PSLG $G = (V, E)$, $E_0 \subseteq E$, with $|E| \leq 2n$ edges.*

Figure 1a, b depicts 1- and 2-regular PSLGs, respectively, where all but one of the vertices are on the boundary of the convex hull. Clearly, the only 3-connected augmentation is the wheel graph, which has $2n-2$ edges. We conjecture that Theorems 1 and 2 can be generalized to PSLGs with maximum degree at most 2.

Conjecture 1.1. Every 3-augmentable PSLG $G_0 = (V, E)$ with $n \geq 4$ vertices and maximum degree at most 2 can be augmented to a 3-connected PSLG $G = (V, E)$, $E_0 \subseteq E$, with $|E| \leq 2n-2$ edges.

It is not possible to extend Theorems 1 and 2 to 3-regular PSLGs. For example, if $G_0 = (V, E_0)$ is a collection of nested 4-cliques as in Fig. 1c, then every 3-connected augmentation requires $3(\frac{n}{4}-1)$ new edges, which gives a total of $\frac{9}{4}n-3$ edges.

As mentioned above, every set of $n \geq 4$ points in the plane $h \leq n$ that lies on the boundary of the convex hull admits a 3-connected PSLG with at most

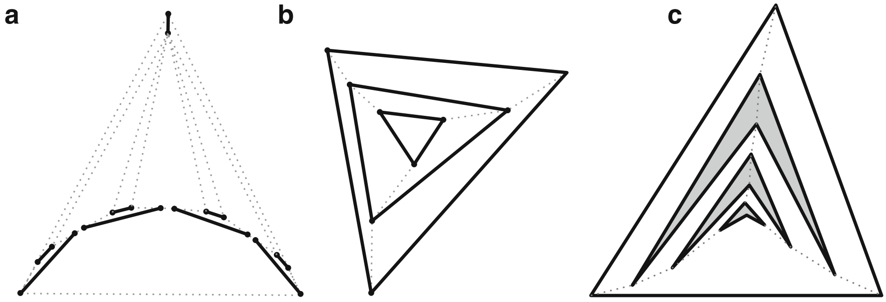


Fig. 2 (**a**) A 1-regular PSLG on n vertices with a triangular convex hull whose 3-connected augmentations have at least $\frac{7}{4}(n-2)$ edges. (**b**) A 2-regular PSLG on n vertices with a triangular convex hull such that every 3-connected augmentation has at least $2n-3$ edges. (**c**) Interior-disjoint simple polygons in a triangular container, for which every 3-connected augmentation has $2n-3$ edges

$\max(\lceil 3n/2 \rceil, n+h-1) \leq 2n-2$ edges [4]. We could not strengthen our Theorems 1 and 2 to be sensitive to the number of hull vertices. Some improvement may be possible for 1-regular PSLGs with fewer than $n-1$ vertices on the convex hull; the best lower bound construction we found with a triangular convex hull requires only $\frac{7}{4}(n-2)$ edges in total (Fig. 2a). For 2-regular PSLGs, however, one cannot expect significant improvement even if $h=3$. If $G_0 = (V, E_0)$ consists of $\frac{n}{3}$ nested triangles (Fig. 2b), then any augmentation to a 3-connected PSLG has at least $2n-3$ edges.

Obstacles in a Container. We have considered whether Theorem 2 can be improved for collections of simple polygons, where the convex hull is a triangle, and there is no nesting among the remaining polygons. We model such 2-regular PSLGs as a collection of *interior-disjoint simple polygons in a triangular container*. Figure 2c shows a construction where every 3-connected augmentation still requires $2n-3$ edges. In this example the polygons are nonconvex, and they are "nested" in the sense that each polygon is visible from at most one larger polygon.

In Sect. 5, we derive lower bounds for the 3-connectivity augmentation of 2-regular PSLGs G_0, where G_0 is a collection of interior-disjoint *convex* polygons (called *obstacles*) lying in a triangular container. All our lower bounds in this section are below $2n-2$, which suggests that Theorem 2 may be improved in this special case.

Organization. In Sect. 2, we introduce a general framework for 3-connectivity augmentation, and prove that every nonconvex simple polygon with n vertices can be augmented to a 3-connected PSLG that has at most $2n-2$ edges. We prove Theorems 1 and 2 in Sects. 3 and 4, respectively. Lower bounds for the model of disjoint convex obstacles in a triangular container are presented in Sect. 5. We conclude with open problems in Sect. 6.

2 Preliminaries

In this section, we prove two preliminary results about *abstract* graphs, which are directly applicable to the 3-connectivity augmentation of simple polygons. In an (abstract) graph $G = (V,E)$, a subset $U \subseteq V$ is called *3-linked* if G contains at least three independent paths between any two vertices of U. (Two paths are *independent* if they do not share any edges or vertices apart from their endpoints.) By Menger's theorem, a graph $G = (V,E)$ is 3-connected if and only if V is 3-linked in G. The following lemma gives a criterion for incrementing a 3-linked set of vertices with one new vertex.

Lemma 2.1. *Let $G = (V,E)$ be a graph such that $U \subset V$ is 3-linked. If G contains three independent paths from $v \in V \setminus U$ to three distinct vertices in U, then $U \cup \{v\}$ is also 3-linked.*

Proof. Assume that G contains three independent paths from $v \in V \setminus U$ to distinct vertices $u_1, u_2, u_3 \in U$. We need to show that for every $u \in U$, there are three independent paths between v and u. By Menger's theorem, it is enough to show that if we delete any two vertices $w_1, w_2 \in V \setminus \{u, v\}$, the remaining graph $G \setminus \{w_1, w_2\}$ still contains a path between v and u. Since there are three independent paths from v to u_1, u_2, and u_3, the graph $G \setminus \{w_1, w_2\}$ contains a path from v to u_i for some $i \in \{1,2,3\}$. If $u_i = u$, then we are done. Otherwise, $G \setminus \{w_1, w_2\}$ contains a path from u_i to u, since U is 3-linked. The union of these two paths (from v to u_i and from u_i to u) contains a path from v to u. □

Lemma 2.2. *Let $G_A = (V,A)$ be a 2-connected graph, and let $G_C = (V,C)$ be a 3-connected graph with $A \subseteq C$. Let $U_A \subseteq V$ be the set of vertices that have degree 3 or higher in G_A, and assume that U_A is 3-linked in G_A. Then $G_A = (V,A)$ can be augmented to a 3-connected graph $G_B = (V,B)$ with $A \subseteq B \subseteq C$, by adding at most $|V \setminus U_A|$ new edges. Furthermore, if $U_A = \emptyset$, then $|V| - 2$ new edges are sufficient for the augmentation.*

Proof. We describe an algorithm that augments $G_A = (V,A)$ to a 3-connected graph $G_B = (V,B)$, $A \subseteq B \subseteq C$. We maintain a graph $G_i = (V,E_i)$ with $A \subseteq E_i \subseteq C$. Initially, we start with $i = 0$ and $E_0 = A$. We augment G_i incrementally by adding new edges from C until G_i becomes 3-connected, and then output $G_B = G_i$. We also increment the set $U_i \subseteq V$ of vertices that have degree 3 or higher in G_i, and maintain the property that U_i is 3-linked in G_i. In each step, we will increment G_i with one new edge such that U_i increases with at least one new vertex (since the endpoints of the new edge will have degree 3 or higher). If $U_i = \emptyset$ or $|U_i| = 2$, then U_i will increase with *two* new vertices in a single step. Our algorithm terminates with $U_i = V$, and the above properties guarantee that altogether at most $|V \setminus U_0|$ new edges are added, and if $U_0 = \emptyset$, then at most $|V| - 2$ new edges are added.

It remains to describe one step of the augmentation, in which we increment E_i with one new edge from C. We distinguish among three cases.

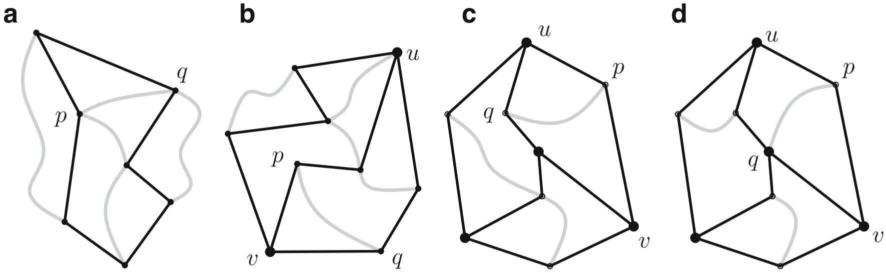


Fig. 3 Illustration for the proof of Lemma 2.2. Edges of G_A are *black*, additional edges of G_C are *gray*, and vertices in U_i are marked with *large dots*. (**a**) G_A is a Hamiltonian cycle. (**b**) G_A has two vertices of degree 3. (**c**) Both p and q lie in the interior of some paths between vertices of U_i. (**d**) Vertex q is in U_i

Case 1. $U_i = \emptyset$. Since G_i is 2-connected, all vertices have degree 2 in G_i, and so it is a Hamiltonian cycle (Fig. 3a). Pick an arbitrary edge $pq \in C \setminus E_i$, and set $E_{i+1} = E_i \cup \{pq\}$. Let $U_{i+1} = \{p,q\}$ be the set of the two vertices of degree 3. Note that U_{i+1} is indeed 3-linked in G_{i+1}, as required.

Case 2. $|U_i| = 2$. Denote the vertices in U_i by u and v. Every edge in E_i is part of a path between u and v (Fig. 3b). Let $\mathcal{P}_i$ denote the set of all (at least three) paths of G_i between u and v. Note that every vertex in $V \setminus U_i$ lies in the interior of a path in $\mathcal{P}_i$. Since G_i is a simple graph, at least two paths in $\mathcal{P}_i$ have interior vertices. Let $P \in \mathcal{P}_i$ be a path with at least one interior vertex. Graph G_C contains some edge $pq \in C$ between an interior vertex p of P and a vertex q outside P; otherwise, the deletion of u and v would disconnect G_C. Set $E_{i+1} = E_i \cup \{pq\}$ and $U_{i+1} = U_i \cup \{p,q\}$. Note that G_{i+1} now contains three independent paths between any two vertices of $U_{i+1} = \{u,v,p,q\}$.

Case 3. $|U_i| \geq 3$. In this case, every edge in E_i is part of a path between two vertices in U_i. Let $\mathcal{P}_i$ denote the set of all paths of G_i between vertices in U_i with no interior vertices in U_i. Note that every vertex in $V \setminus U_i$ lies in the interior of a path in $\mathcal{P}_i$. Pick two vertices $u, v \in U_i$ connected by a path in $\mathcal{P}_i$ with at least one interior vertex, and let V_{uv} be the set of interior vertices of all paths in $\mathcal{P}_i$ between u and v (Fig. 3c, d). Graph G_C contains some edge $pq \in C$ between a vertex $p \in V_{uv}$ and a vertex q outside $V_{uv} \cup \{u,v\}$; otherwise, the deletion of u and v would disconnect G_C. Set $E_{i+1} = E_i \cup \{pq\}$. Now G_{i+1} contains three independent paths from p to three vertices of U_i: independent paths to u and v along a path in $\mathcal{P}_i$, and a third path starting with edge pq and if $q \notin U_i$, then continuing along a path containing q to a third vertex in $U_i \setminus \{u,v\}$. Similarly, if $q \notin U_i$, then G_{i+1} contains three independent paths from q to three vertices of $U_i \cup \{p\}$. By Lemma 2.1, $U_i \cup \{p,q\}$ is 3-linked in G_{i+1}. So we can set $U_{i+1} = U_i \cup \{p,q\}$. □

We show next that every simple polygon in the plane with $n \geq 3$ vertices can be augmented to a 3-connected PSLG that has at most $2n-2$ edges.

Corollary 2.3. *Every simple polygon with $n \geq 4$ vertices, not all in convex position, can be augmented to a 3-connected PSLG that has at most $2n-2$ edges.*

Proof. The edges and vertices of a simple polygon form a Hamiltonian cycle $G_A = (V,A)$. By the results of Valtr and Tóth [13], if the polygon is nonconvex, then it is 3-augmentable, so there is a 3-connected PSLG $G_C = (V,C)$, $A \subset C$. Lemma 2.2 completes the proof. □

3 Disjoint Line Segments

In this section, we prove Theorem 1. Let $G_A = (V,A)$ be a straight-line embedding of a perfect matching with $n \geq 4$ vertices, not all in convex position. We show that if no edge in A is a proper chord of the convex hull of the vertices, then G_A can be augmented to a 3-connected PSLG that has at most $2n-2$ edges. We use the result by Hoffmann and Tóth [5] that G_A can be augmented to a Hamiltonian PSLG G_H. If G_A is 3-augmentable, then G_H is also 3-augmentable and can be augmented to a 3-connected Hamiltonian PSLG G_C. In the following lemma, we use such a graph G_C, but we no longer rely on its *straight-line* embedding; our argument works for any plane drawing (where the edges are represented by Jordan arcs).

Lemma 3.1. *Let $G_A = (V,A)$ be a perfect matching with $n \geq 4$ vertices, and let $G_C = (V,C)$ be a 3-connected Hamiltonian planar graph with $A \subseteq C$. Then there is a 3-connected graph $G_B = (V,B)$ such that $A \subseteq B \subseteq C$ and $|B| \leq 2n-2$.*

Proof. Fix an arbitrary plane embedding of G_C. Let (V,H) be an arbitrary Hamiltonian cycle in G_C. If $A \subset H$, then the result follows from Lemma 2.2. Assume that $A \not\subset H$.

We construct a 3-connected graph $G_B = (V,B)$, $A \subseteq B \subseteq C$, incrementally. We maintain a 2-connected graph $G_i = (V,E_i)$ with $E_i \subseteq C$ (but E_i does not necessarily contain A). Let $U_i \subseteq V$ denote the set of *all* vertices that have degree at least 3 in G_i, which are called *hubs*. The hubs decompose G_i into a set $\mathcal{P}_i$ of paths between hubs. We maintain the following invariants for G_i.

- **I**$_1$ $U_i \subseteq V$ is 3-linked in G_i.
- **I**$_2$ Between any two hubs in U_i, there are at most two paths in $\mathcal{P}_i$, at most one of which has interior vertices.
- **I**$_3$ If there is an edge $uv \in A \setminus E_i$ between two hubs $u,v \in U_i$, then there is also a path in $\mathcal{P}_i$ between u and v (such a path is called a lens).
- **I**$_4$ $|E_i| \leq (n-2) + |U_i| - b_i$, where b_i is the number of bad paths in $\mathcal{P}_i$ (defined below).

In the next paragraphs, we define three families of so-called bad paths in $\mathcal{P}_i$ (lenses, diamonds, and monsters). Some of the definitions are formulated for subpaths of a path $P \in \mathcal{P}_i$ in order to keep track of subpaths that may become a bad path in $\mathcal{P}_{i+1}$ when some interior vertices of P become hubs. We begin by introducing some notation for paths. For two vertices, p and q, of a path $P \in \mathcal{P}_i$, let

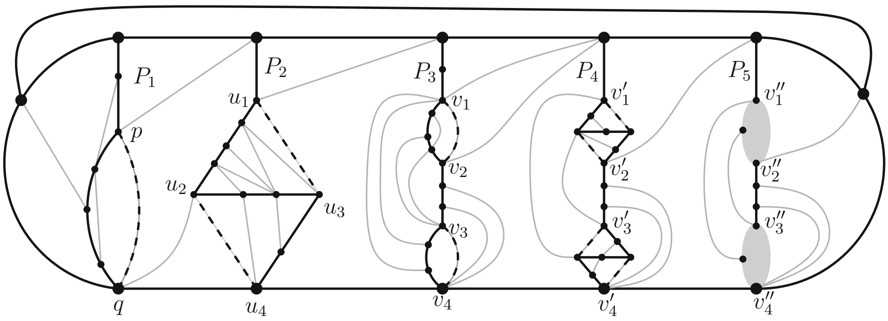


Fig. 4 A lens $P_1[p,q]$. A diamond path $P_2[u_1,u_4]$. Monsters $P_3[v_1,v_4]$, $P_4[v'_1,v'_4]$, and $P_5[v''_1,v''_4]$. The vertices p, u_1, v_1, v'_1, and v''_1 are dangerous. *Large dots* are hubs in U_i, *solid edges* are in E_i, *dashed edges* are in $A \setminus E_i$, and *gray edges* are in $C \setminus E_i$, respectively. *Gray ovals* represent a lens, a diamond, or a monster

$P[p,q]$ denote the subpath of P between p and q (if p and q are the two endpoints of P, then $P = P[p,q]$). We say that the edges between interior vertices of $P[p,q]$ and vertices outside $P[p,q]$ *go out of* $P[p,q]$.

Let P be a path in $\mathcal{P}_i$. Refer to Fig. 4. A subpath $P[u,v] \subseteq P$ is a *lens* if $uv \in A \setminus E_i$. A subpath $P[u_1,u_4] \subseteq P$ is a *diamond* if there are vertices u_1,u_2,u_3,u_4 along P in this order such that

(1) $u_1u_3, u_2u_4 \in A \setminus E_i$, and
(2) Every edge going out of $P[u_1,u_4]$ is incident to u_2 or u_3.

The third family of subpaths, called monsters, is defined recursively. A subpath $P[v_1,v_4] \subseteq P$ is a *monster* if there are vertices v_1,v_2,v_3,v_4 along P in this order such that

(1) Each of $P[v_1,v_2]$ and $P[v_3,v_4]$ is a lens, a diamond, or a smaller monster.
(2) $P[v_2,v_3]$ has at least one interior vertex.
(3) Every edge going out of $P[v_1,v_4]$ is incident to v_2.
(4) Every edge going out of $P[v_1,v_2]$ is incident to v_3, every edge going out of $P[v_2,v_3]$ is incident to v_4, and every edge going out of $P[v_3,v_4]$ is incident to v_1.

In a minimal monster, each of $P[v_1,v_2]$ and $P[v_3,v_4]$ is either a lens or a diamond. We say that a (sub)path $P' \subseteq P \in \mathcal{P}_i$ is *dangerous* if it is a lens, a diamond, or a monster. Note that every dangerous path has at least one interior vertex. A key property of a dangerous path P' is that each endpoint of P' is incident to some edge in $A \setminus E_i$ that goes to some other vertex of P'. For example, this implies that the middle portion $P[v_2,v_3]$ of a monster is *not* dangerous. A vertex p is called *dangerous* if p is an interior vertex of a path $P \in \mathcal{P}_i$ with endpoints u and v, and either $P[p,u]$ or $P[p,v]$ is a dangerous path. Note that at most one of $P[p,u]$ and $P[p,v]$ can be dangerous, since p is incident to at most one edge of $A \setminus E_i$.

We are now ready to define bad and good paths in $\mathcal{P}_i$. A path $P \in \mathcal{P}_i$ is *bad* if it is dangerous, and *good* otherwise. We denote by b_i the number of bad paths in $\mathcal{P}_i$.

In each step of our algorithm, we modify E_i so that the set of hubs U_i strictly increases and invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained. The algorithm terminates when $U_i = V$. At that time, G_i is a 3-connected subgraph of G_C by invariant $\mathbf{I}_1$, every path in $\mathcal{P}_i$ is a single edge (hence there is no bad path in $\mathcal{P}_i$), all edges of A are contained in E_i by invariant $\mathbf{I}_3$, and $|E_i| \leq 2n-2$ by invariant $\mathbf{I}_4$. So we can output $G_B = G_i$. We note here that the set of edges, E_i, does not always increase. Sometimes we may delete an edge from E_i (and add several edges from $C \setminus E_i$) to obtain E_{i+1}.

Initialization. Recall that (V,H) is a Hamiltonian cycle in G_C, with n edges, such that $A \not\subset H$. Let $pq \in A$ be an arbitrary chord of H. Vertices p and q decompose the Hamiltonian cycle into two paths, each of which has some interior vertices. Since G_C is 3-connected, it contains an edge $st \in C$ between two interior vertices of the two paths. Let $G_0 = (V,E_0)$ with $E_0 = H \cup \{pq, st\}$. Then the set of hubs is $U_0 = \{p,q,s,t\}$, which is 3-linked in G_0. The matching A contains pq and possibly st, so E_0 contains all edges of A induced by U_0. There are six paths in $\mathcal{P}_0$ between hubs. Two of these paths, pq and st, have no interior vertices; hence they are good. The other four paths are incident to p or q, where the incident edge of the matching is $pq \in A$, and so these paths are good as well. We have $|E_0| = n+2$, $|U_0| = 4$, and $b_0 = 0$, which gives $|E_0| = (n-2) + |U_0| - b_0$. The initial graph G_0 satisfies invariants $\mathbf{I}_1$–$\mathbf{I}_4$.

General Step *i*. We are given a graph $G_i = (V,E_i)$ satisfying invariants $\mathbf{I}_1$–$\mathbf{I}_4$ and $U_i \neq V$. We construct a graph G_{i+1} maintaining invariants $\mathbf{I}_1$–$\mathbf{I}_4$ so that the set of hubs strictly increases. Let X_i be the set of vertices $x \in V \setminus U_i$ such that x is an interior vertex of some path $P_x \in \mathcal{P}_i$, and it is adjacent to a vertex outside P_x.

Remark 3.2. In all cases discussed below, we augment G_i with an edge xy, where $x \in X_i$ is an interior vertex of a path $P_x \in \mathcal{P}_i$ and y is outside path P_x. Vertex y is either a hub already in U_i or an interior vertex of some path $P_y \in \mathcal{P}_i$. In the latter case, y becomes a hub with three independent paths to to three hubs in U_i by invariants $\mathbf{I}_1$ and $\mathbf{I}_2$. It decomposes P_y into two paths satisfying $\mathbf{I}_2$ and $\mathbf{I}_3$, and at most one of them can be dangerous. One additional hub at y and one possible new bad path cannot decrease the right-hand side of inequality $|E_i| \leq (n-2) + |U_i| - b_i$ in Invariant $\mathbf{I}_4$. For verifying that invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained, we may assume that y is already a hub in U_i, unless stated otherwise. In all cases below we focus on x and the path $P_x \in \mathcal{P}_i$.

Dispersing Dangerous Paths. We first show that if there is a bad path $P \in \mathcal{P}_i$, then we can always augment G_i so that the new hubs break P into good paths. The augmentation operation `Disperse` below is formulated in a more general setting, since it is a basic building block in several cases below. In certain cases, it is applied for a graph that satisfies invariants $\mathbf{I}_1$–$\mathbf{I}_3$ only.

`Disperse` (G_i, P', xy).
Input: $G_i = (V,E_i)$ is graph satisfying invariants $\mathbf{I}_1$–$\mathbf{I}_3$, P' is a dangerous subpath of some path $P \in \mathcal{P}_i$ (possibly, $P' = P_x$), and $xy \in C$ is an edge between a vertex $x \in X_i$ in the interior of P' and a vertex outside P, such that either $xy \in A \setminus E_i$ or none of the edges incident to x and going out of P' is in $A \setminus E_i$.

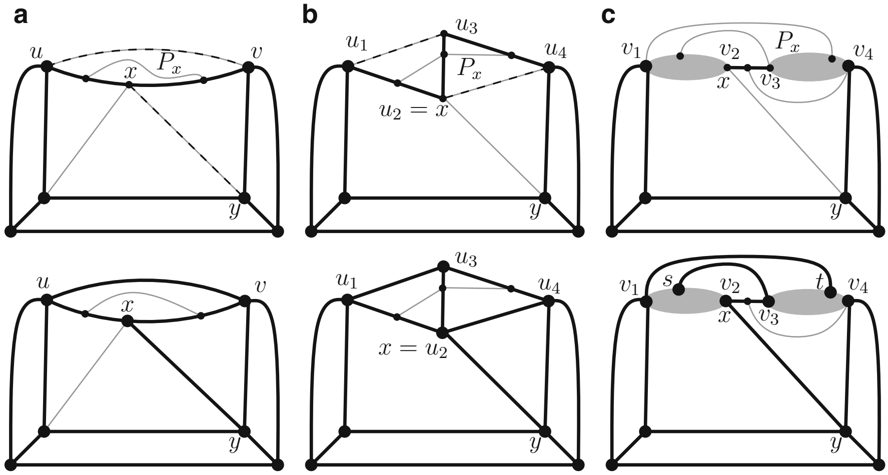


Fig. 5 Operation $\texttt{Disperse}(G_i, P', xy)$ where $P' = P$ and P is a lens (**a**), a diamond (**b**), and a monster (**c**). The *top row* shows G_i, and the *bottom row* G_{i+1}. Vertex x is in the interior of a dangerous path $P' \subseteq P$ and y is a hub outside path P. *Solid edges* are in E_i, *dashed edges* are in $A \setminus E_i$, and *dotted edges* are in $C \setminus (A \cup E_i)$

1. If P' is a lens with endpoints u and v, then $E_i := E_i \cup \{xy, uv\}$ (see Fig. 5a).
2. If P' is a diamond defined by vertices (u_1, u_2, u_3, u_4) such that $x = u_2$, then $E_i := E_{i+1} \cup \{xy, u_1u_3, u_2u_4\}$ (see Fig. 5b).
3. If P' is a monster defined by vertices (v_1, v_2, v_3, v_4), then augment G_i in three steps (Fig. 5c): set $E_i := E_i \cup \{xy\}$; call $\texttt{Disperse}(G_i, P[v_1, v_2], sv_3)$ for an edge sv_3 going out of $P[v_1, v_2]$; and call $\texttt{Disperse}(G_i, P[v_3, v_4], tv_1)$ for an edge tv_1 going out of $P[v_3, v_4]$.

Let $E_{i+1} = E_i$ and return $G_{i+1} = (V, E_{i+1})$.

We show that operation Disperse maintains invariants $\mathbf{I}_1$–$\mathbf{I}_3$. Instead of maintaining invariant $\mathbf{I}_4$, and we show that the value of $|E_i| - |U_i| + b_i$ does not increase.

Proposition 3.3. *Operation* $\texttt{Disperse}(G_i, P', xy)$ *augments* G_i *to* G_{i+1} *such that*

- *Invariants* $\mathbf{I}_1$–$\mathbf{I}_3$ *are maintained;*
- $|E_{i+1}| - |U_{i+1}| + b_{i+1} \le |E_i| - |U_i| + b_i$*;*
- *the subgraph of* G_{i+1} *induced by the vertices of* P' *contain a simple cycle* σ_x *passing through vertex* x *and the two endpoints of* P'*.*

The second condition implies, in particular, that invariant $\mathbf{I}_4$ is maintained if G_i satisfies $\mathbf{I}_4$.

Proof. We proceed by induction on the length of P'. We distinguish three cases, depending upon whether P' is a lens, a diamond, or a monster. We may assume $y \in U_i$ by Remark 3.2.

Case (α): P' is a lens. In this case, $U_{i+1} = U_i \cup \{x,u,v\}$. The vertices of P' induce a cycle $\sigma_x = P' \cup \{uv\}$ in G_{i+1}. The vertices u, v, and x each have three independent paths to the two endpoints of P and to y in G_{i+1}. By invariants $\mathbf{I}_1$, $\mathbf{I}_2$, and Lemma 2.1, U_{i+1} is 3-linked in G_{i+1}. It is easy to verify that invariants $\mathbf{I}_2$–$\mathbf{I}_3$ are also maintained. All new paths in $\mathcal{P}_{i+1}$ are good. We have added two new edges. Vertex x is always a new hub. If x is the only new hub, then $P' = P_x$ and so $b_{i+1} = b_i - 1$. At any rate, we have $|E_{i+1}| - |U_{i+1}| + b_{i+1} \leq |E_i| - |U_i| + b_i$.

Case (β): P' is a diamond. In this case, $U_{i+1} = U_i \cup \{u_1,u_2,u_3,u_3\}$. The vertices u_1,u_2,u_3,u_4 are in the simple cycle $\sigma_x = P[u_1,u_2] \cup \{u_2u_4\} \cup P[u_4,u_3] \cup \{u_3u_1\}$ in G_{i+1}. Each new hub has three independent paths to the two endpoints of P and to y. By $\mathbf{I}_1$, $\mathbf{I}_2$, and Lemma 2.1, U_{i+1} is 3-linked in G_{i+1}. It is easy to verify that invariants $\mathbf{I}_2$–$\mathbf{I}_3$ are also maintained. All new paths in $\mathcal{P}_{i+1}$ are good. We have added three new edges. Vertices u_2,u_3 are always new hubs. If u_2 and u_3 are the only new hubs, then $P' = P$ and $b_{i+1} = b_i - 1$. Hence, we have $|E_{i+1}| - |U_{i+1}| + b_{i+1} \leq |E_i| - |U_i| + b_i$.

Case (γ): P' is a monster. It is easy to verify that the first step maintains $\mathbf{I}_1$–$\mathbf{I}_3$, and the last two steps maintain $\mathbf{I}_1$–$\mathbf{I}_3$ by induction. We construct a simple cycle σ_x in the subgraph of G_{i+1} induced by the vertices of P' that passes through v_1, v_2, and v_4. We construct σ_x explicitly as a union of two independent arcs between v_1 and v_3, one of which passes through $v_2 = x$, the other one through v_4. Refer to Fig. 5c. In the second step, we added an edge sv_3 for some interior vertex s of path $P[v_1,v_2]$, and there is a simple cycle σ_s through v_1, v_2, and s by indiction. Let one arc of σ_x be the union of v_3s and the part of σ_s from s to v_1 passing through v_2. In the third step, we added an edge tv_1 for some interior vertex t of path $P[v_3,v_4]$, and there is a simple cycle σ_t through v_3, v_4, and t by induction. Let the second arc of σ_x be the union of v_1t and the part of C_t from t to v_3 passing through v_4.

To verify $|E_{i+1}| - |U_{i+1}| + b_{i+1} \leq |E_i| - |U_i| + b_i$, we distinguish two subcases.

Case ($\gamma1$): $P' = P$. If $P' = P$, then the first two steps preserve the number of bad paths (P is bad initially, $P[v_1,v_2]$ is bad after the first step, and $P_x[v_3,v_4]$ is bad after the second step); however, all new paths in $\mathcal{P}_{i+1}$ are good. Altogether, we have $b_{i+1} < b_i$. The quantity $|E_i| + b_i - |U_i|$ is unchanged in the first step, and it can only decrease in the last two steps by induction.

Case ($\gamma2$): $P' \neq P$. If $P' \neq P$, then at least one endpoint of P' is an interior vertex of P and becomes a hub in G_{i+1}. Path $P \in \mathcal{P}_i$ may be either good or bad, and all new paths in $\mathcal{P}_{i+1}$ are good. Compared to case 3(a), the number of hubs increases by one, while the first step may increase the number of bad paths; that is, $b_{i+1} \leq b_i + 1$. Altogether, invariant $\mathbf{I}_4$ is maintained. □

Case Analysis. We are given a graph $G_i = (V,E_i)$ satisfying invariants $\mathbf{I}_1$–$\mathbf{I}_4$ and $U_i \neq V$, and we would like to construct a graph G_{i+1} maintaining invariants $\mathbf{I}_1$–$\mathbf{I}_4$ such that $|U_i| < |U_{i+1}|$. We proceed with a case analysis, distinguishing among three main cases.

Case 1. **There is a vertex $x \in X_i$ that is not dangerous.** Refer to Fig. 6a. If there is an edge in $A \setminus E_i$ between x and a vertex outside P_x, then let this edge be xy;

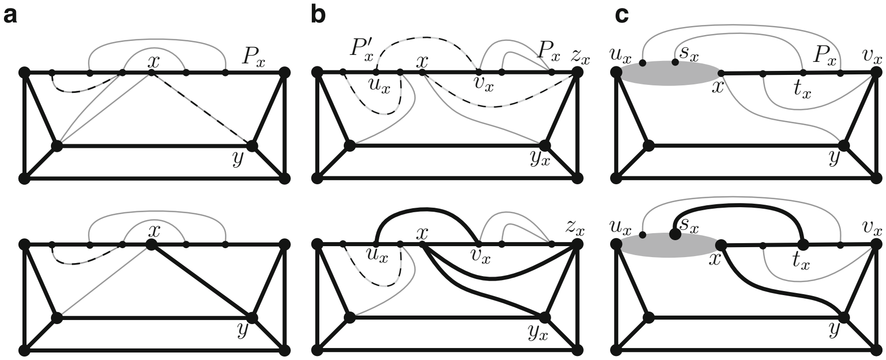


Fig. 6 Step i of the algorithm. The *top row* shows G_i, and the *bottom row* G_{i+1}. (**a**) Case 1: vertex $x \in X_i$ is not dangerous. (**b**) Case 2: $x \in X_i$ is an interior vertex of a dangerous subpath $P'_x \subseteq P_x$, and it is incident to an edge $xz_x \in A \setminus E_i$. (**c**) Case 3(a): $x \in X_i$ is a dangerous vertex of P_x because subpath $P_x[u_x, x] \subset P_x$ is dangerous, but t_x is not a dangerous vertex

otherwise, let xy be an arbitrary edge between x and a vertex outside P_x. Let $E_{i+1} = E_i \cup \{xy\}$. For verifying that invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained, we may assume $y \in U_i$ by Remark 3.2. Then we have $U_{i+1} = U_i \cup \{x\}$, and U_{i+1} is 3-linked in G_{i+1} by $\mathbf{I}_1$, $\mathbf{I}_2$, and Lemma 2.1. The new hub x subdivides P_x into two good paths in $\mathcal{P}_{i+1}$ (even if $P_x \in \mathcal{P}_i$ is bad). It is easily checked that invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained.

Case 2. **There is a dangerous vertex $x \in X_i$ such that x is in the interior of a dangerous path P'_x, $P'_x \subseteq P_x \in \mathcal{P}_i$.** Let $\widehat{X}_i \subseteq X_i$ be the set of all vertices $x \in X_i$ that are dangerous and lie in the interior of some dangerous subpath $P'_x \subset P_x \in \mathcal{P}_i$.

For every $x \in \widehat{X}_i$, let y_x be an arbitrary edge between x and a vertex y_x outside P_x. Let $P'_x \subseteq P_x$ be a maximal dangerous subpath of P_x that contains x in its interior. If there is a vertex $x \in \widehat{X}_i$ such that $A \setminus E_i$ contains no edge incident to x going out of P'_x, then we apply operation $\texttt{Disperse}(G_i, P'_x, xy)$, and invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained by Proposition 3.3.

Assume now that for every $x \in \widehat{X}_i$, there is an edge $xz_x \in A \setminus E_i$ such that z_x is outside P'_x. Note that z_x is a vertex of P_x; otherwise, x would not be a dangerous vertex. Note that P'_x must be a lens since the edges going out of a diamond or a monster are not in $A \setminus E_i$. Denote the endpoints of P'_x by u_x and v_x, where $u_x v_x \in A \setminus E_i$. The vertices u_x, v_x, x, z_x are pairwise distinct (they are the endpoints of two edges in the matching A). Refer to Fig. 6b.

We show that there is an $x \in \widehat{X}_i$ such that at least three vertices in $\{u_x, v_x, x, y_x\}$ are in the interior of P_x. Suppose, to the contrary, that for every $x \in \widehat{X}_i$, the two endpoints of P_x are in $\{u_x, v_x, x, y_x\}$. Without loss of generality, the endpoints of P_x are u_x and z_x. Then each interior vertex of P_x lies in the interior of lens $P_x[u_x, v_x]$ or lens $P_x[x, z_x]$. So every edge going out of P_x is incident to a vertex in $\widehat{X}_i$. Moreover, every $x \in \widehat{X}_i$ is joined to an endpoint of P_x. Therefore, P_x is a diamond for every $x \in \widehat{X}_i$, and the maximum dangerous subpath containing x in its interior is $P'_x = P_x$, contradicting our assumption that $P'_x = P_x[u_x, v_x]$.

Let $x \in \widehat{X}_i$ such that at least three vertices in $\{u_x, v_x, x, y_x\}$ are in the interior of P_x. Set $E_{i+1} = E_i \cup \{u_x v_x, x y_x, x z_x\}$. For the verifying invariants $\mathbf{I}_1$–$\mathbf{I}_4$ for G_{i+1}, we may assume $y \in U_i$ by Remark 3.2. Then $U_{i+1} = U_i \cup \{u_x, v_x, x, z_x\}$. We have added three new edges and at least three new hubs. The new hubs subdivide P_x into good paths, and so we have $b_{i+1} \leq b_i$. It is easily checked that invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained.

Case 3. **Every $x \in X_i$ is a dangerous vertex of P_x but not an interior vertex of any dangerous subpath of P_x.** We introduce some notation and then distinguish among four subcases. For every $x \in X_i$, denote the endpoints of P_x by $u_x, v_x \in U_i$. Assume, without loss of generality, that the subpath $P_x[u_x, x]$ of P_x is dangerous (hence $P_x[v_x, x]$ is not dangerous).

There is no vertex $x' \in X_i$ in the interior of $P_x[u_x, x]$ since $P_x[u_x, x]$ is a dangerous subpath of P_x. It follows that every path $P \in \mathcal{P}_i$ contains at most two vertices from X_i. Moreover, if $x, x' \in X_i$ are two distinct interior vertices of some path $P \in \mathcal{P}_i$, then the two endpoints of P are u_x and $u_{x'}$, and the subpaths $P[u_x, x]$ and $P[u_{x'}, x']$ are disjoint.

For every $x \in X_i$, the subpath $P_x[u_x, x]$ has at least one interior vertex (because $P_x[u_x, x]$ is dangerous). Since G_C is 3-connected, there must be at least one edge going out of $P_x[u_x, x]$. Let C_x denote the set of edges in C going out of $P_x[u_x, x]$. As noted above, all edges in C_x are incident to some vertex of P_x outside $P_x[u_x, x]$. For every edge $s_x t_x \in C_x$, we use the convention that s_x denotes an interior vertex of $P_x[u_x, x]$ and t_x denotes a vertex of P_x outside $P_x[u_x, x]$.

Note that if an edge $s_x t_x \in C_x$ is in $A \setminus E_i$, then t_x is not a dangerous vertex. Indeed, if t_x is dangerous, then either $P_x[u_x, t_x]$ or $P_x[v_x, t_x]$ is a dangerous subpath. However, $P_x[u_x, t_x]$ cannot be dangerous, since it contains x in its interior, and we assumed that x is not in the interior of any dangerous subpath of P_x. Hence, $P_x[v_x, t_x]$ is dangerous, and so an edge in $A \setminus E_i$ joins t_x to some other vertex in $P_x[v_x, t_x]$.

Case 3(a): There are $x \in X_i$ and $s_x t_x \in C_x$ such that t_x is not a dangerous vertex. If C_x contains an edge in $A \setminus E_i$, then let $s_x t_x$ be such an edge; otherwise, let $s_x t_x \in C_x$ be an arbitrary edge such that t_x is not dangerous. Refer to Fig. 6c. Construct E_{i+1} from E_i in two steps as follows: Set $E_i := E_i \cup \{xy\}$, and call `Disperse`$(G_i, P_x[u_x, x], s_x t_x)$. It is easily checked that the first step maintains $\mathbf{I}_1$–$\mathbf{I}_4$, and the second step maintains $\mathbf{I}_1$–$\mathbf{I}_4$ by Proposition 3.3.

In the remaining cases [Cases 3(b)–(d)], we assume that for every $x \in X_i$, the edges $s_x t_x \in C_x$ are not in $A \setminus E_i$; otherwise, Case 3(a) would apply.

Case 3(b): There are $x \in X_i$ and $s_x t_x \in C_x$ such that t_x is in the interior of a dangerous subpath P'' of $P_x[x, v_x]$. Refer to Fig. 7a. Construct E_{i+1} from E_i in two steps as follows: Set $E_i := E_i \cup \{xy\}$, and then apply independently `Disperse`$(G_i, P_x[u_x, x], s_x t_x)$, and `Disperse`$(G_i, P'', s_x t_x)$. The first step increases $|E_i| - |U_i| + b_i$. However, the two independent calls to `Disperse` add the same edge $s_x t_x$, and so $|E_i| - |U_i| + b_i$ decreases by at least one. Altogether, invariants $\mathbf{I}_1$–$\mathbf{I}_4$ are maintained.

Case 3(c): There are $x \in X_i$ and $s_x t_x \in C_x$ such that t_x is a dangerous vertex but $P_x[x, t_x]$ has no interior vertices. Refer to Fig. 7b. Construct E_{i+1} from E_i in three

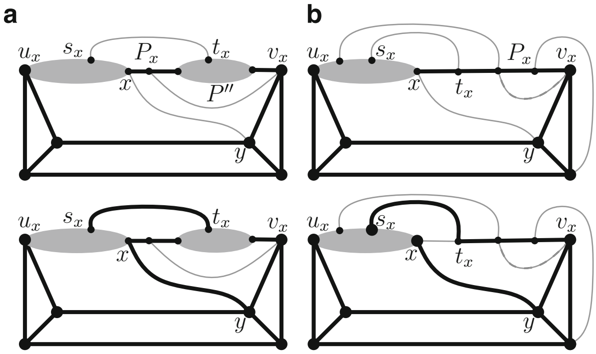


Fig. 7 Step i of the algorithm. The *top row* shows G_i, and the *bottom row* G_{i+1}. (**a**) Case 3b: subpaths $P_x[u_x,x] \subset P_x$ and $P_x[v_x,t_x]$ are dangerous and t_x is in the interior of $P_x[v_x,t_x]$. (**b**) Case 3c: subpath $P_x[u_x,x] \subset P_x$ is dangerous, but $P_x[x,t_x]$ has no interior vertices. (**c**) Case 3d: t_x is not adjacent to any vertex outside P_x. There is an edge $s't'$ where s' is an interior vertex of $P_x[v_x,t_x]$ and t' is an interior vertex of $P_x[x,t_x]$

steps as follows: Set $E_i := E_i \cup \{xy\}$, call operation `Disperse` for path $P_x[u_x,x]$ and edge s_xt_x, and then delete edge xt_x. The first two steps clearly maintain invariants $\mathbf{I}_1$–$\mathbf{I}_3$, but we need to be careful about the edge deletion. We show that after the deletion of xt_x, the original hubs in U_i remain 3-linked, vertex x becomes a hub with three independent paths to hubs in U_i; and t_x will be a hub in G_{i+1} if and only if it was already a hub in G_i. Indeed, the two endpoints of P_x are connected by $P_x[u_x,s_x] \cup \{s_xt_x\}$, so U_i remains 3-linked. In the second step, we create a simple cycle σ_{s_x} passing through x, s_x, and u_x by Proposition 3.3. Hence, the degree of x is at least 3 in G_{i+1}, and it has three independent paths to u_x, v_x, and y. Finally, the degree of t_x does not change, and so it is a hub if and only if $t_x \in U_i$. So the third step also maintains invariants $\mathbf{I}_1$–$\mathbf{I}_3$.

For invariant $\mathbf{I}_4$, notice that the number of dangerous paths can only decrease by Proposition 3.3 (that is $b_{i+1} \le b_i$), and the effects of inserting edge xy and deleting edge xt_x cancel each other. Hence, invariant $\mathbf{I}_4$ is maintained.

Case 3(d): For every $x \in X_i$ and $s_xt_x \in C_x$, vertex t_x is dangerous, it is not in the interior of any dangerous subpath of $P_x[x,v_x]$, and $P_x[x,t_x]$ has some interior vertices. For every $x \in X_i$, we choose an edge $s_xt_x \in C_x$ as follows. Let $s_xt_x \in C_x$ be an edge such that t_x is the closest vertex to x along P_x. Fix a vertex $x \in X_i$ such that the length of $P_x[x,t_x]$ is minimal.

Path P_x with vertices u_x, x, t_x, v_x satisfies conditions (1)–(2) in the definition of monsters. The vertices u_x, x, t_x, v_x appear in this order along P_x such that $P_x[u_x,x]$ and $P_x[t_x,v_x]$ are dangerous and $P_x[x,t_x]$ has at least one interior vertex. We show next that it satisfies condition (3) as well.

We show that every vertex going out of P_x is incident to x. Suppose to the contrary that there is a vertex $x' \in X_i$, $x' \neq x$, in P_x. Then $P_x[x',v_x]$ is a dangerous subpath, which is disjoint from $P_x[u_x,x]$. Vertex $x' \in X_i$ is not in the interior of any dangerous

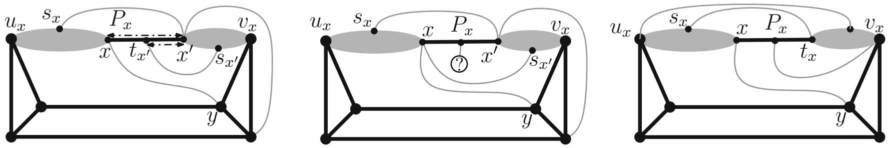


Fig. 8 Case 3(d): subpaths $P_x[u_x, x] \subset P_x$ and $P_x[v_x, t_x]$ are dangerous. Suppose that $t_x = x'$ is adjacent to a vertex outside P_x. *Left*: if $t_{x'}$ is in the interior of $P_x[x, x']$, then $P_x[x', t_{t'}]$ is shorter than $P_x[x, t_x]$. *Middle*: if $t_{x'} = x$, then the are no edges out of $P_x, [x, x']$. *Right*: if there no edge between the interiors of $P_x[x, t_x]$ and $P_x[t_x, v_x]$, then all edges out of $P_x[t_x, v_x]$ go to v_1

subpath of P_x. So x' is in $P_x[x, t_x]$ (possibly, $x' = t_x$). Note that x' is not in the interior of $P_x[x, t_x]$; otherwise, t_x would be in the interior of the dangerous subpath $P_x[x', v_x]$ of $P_x[x, v_x]$. The only remaining possibility is $x' = t_x$. For vertex $x' \in X_i$, we have defined an edge $s_{x'}t_{x'}$, where $s_{x'}$ is an interior vertex of $P_x[x', v_x] = P_x[t_x, v_x]$ and $t_{x'}$ is a vertex of P_x outside $P_x[v_x, x']$. However, $t_{x'}$ cannot be at u_x or in the interior of the dangerous path $P_x[u_x, x]$; otherwise, Case 3(a) or 3(b) would apply for x'. Also, $t_{x'}$ cannot be in the interior of $P_x[x, x']$ because then subpath $P_{x'}[x', t_{x'}]$ would be strictly shorter than $P_x[x, t_x]$, contradicting the choice of $x \in X_i$ (Fig. 8, left). Therefore, we have $t_{x'} = x$ (Fig. 8, middle). Now consider the interior vertices of $P_x[x, x']$. These vertices are adjacent to the interior of neither $P_x[u_x, x]$ nor $P_x[x', v_x]$ by the choice of the edges $s_x t_x$ and $s_{x'} t_{x'}$. They are separated from all other vertices outside $P_x[x, x']$ by the cycle $\{s_x x'\} \cup P_x[x', s_{x'}] \cup \{s_{x'} x\} \cup P_x[x s_x]$. There are no edges going out of $P_x[x, x']$, contradicting the 3-connectivity of G_C. We conclude that every vertex going out of P_x is incident to x.

We also show that every edge going out of $P_x[u_x, x]$ is incident to t_x. Indeed, as noted above, all edges in C_x are incident to a vertex in P_x outside $P_x[u_x, x]$. They cannot be incident to interior vertices of $P_x[x, t_x]$ by the choice of $s_x t_x$. They also cannot be incident to v_x or any interior vertex of $P_x[t_x, v_x]$; otherwise, Case 3(a) or 3(b) would apply.

Next we show that there is an edge between the interior of $P_x[x, t_x]$ and the interior of $P_x[t_x, v_x]$. Suppose to the contrary that there is no such edge. Consider the edges going out of $P_x[x, t_x]$. They are not incident to any vertex outside P_x. By our assumption, they are not incident to any interior vertex of $P_x[t_x, v_x]$. They are not incident to any interior vertex of $P_x[u_x, x]$ by the choice of $s_x t_x$. They are also not incident to u_x because of planarity: Path $P_x[v_x, t_x] \cup \{s_x t_x\} \cup P_x[s_x, x] \cup \{xy\}$ separates them from u_x. Hence, all edges out of $P_x[x, t_x]$ are incident to v_x. Note that $s_x t_x$ and the edges out of $P_x[x, t_x]$ lie on opposite sides of path P_x by planarity. Now consider the edges out of $P_x[t_x, v_x]$. They are not incident to any interior vertex of $P_x[u_x, x]$ as noted above. They are not incident to interior vertices of $P_x[x, t_x]$ by our assumption. They are also not incident to x because $s_x t_x$ and an edge going out of $P_x[x, t_x]$ to v_x separate them from x. Therefore, all edges going out of $P_x[t_x, v_x]$ are incident to u_x. (Fig. 8, right). This means that P_x is a monster, and Case 2 would apply. We conclude that some edge goes out of $P_x[t_x, v_x]$ to an interior vertex of $P_x[x, t_x]$.

Let $s't'$ be an edge between an interior vertex s' of $P_x[t_x, v_x]$ and an interior vertex t' of $P_x[x,t_x]$. If any edge going out of $P_x[x,t_x]$ is in $A \setminus E_i$, then let one of them be $s't'$; otherwise, we can choose $s't'$ arbitrarily. Refer to Fig. 7c. Construct E_{i+1} from E_i in three steps as follows: Set $E_i := E_i \cup \{xy\}$, call operation `Disperse` for path $P_x[u_x,x]$ and edge s_xt_x, and then call `Disperse` for path $P_x[t_x, v_x]$ and edge $s't'$ (which introduces a new hub at t'). Invariants $\mathbf{I}_1$–$\mathbf{I}_3$ are maintained in all three steps by Proposition 3.3. For invariant $\mathbf{I}_4$, the first step increases $|E_i| - |U_i| + b_i$ by one, the second step maintains it by Proposition 3.3, and the third step decreases it because of the extra hub at t'. Altogether, invariant $\mathbf{I}_4$ is maintained. This completes the description of Case 3(d).

While $U_i \neq V$, we can apply Case 1, 2, or 3 and increase the number of hubs. If $U_i = V$, then $G_i = (V, G_i)$ is a 3-connected graph such that $A \subseteq E_i \subseteq C$, as discussed above. This completes the proof of Lemma 3.1. □

Corollary 3.4. *Every 3-augmentable planar straight-line matching with $n \geq 4$ vertices can be augmented to a 3-connected PSLG that has at most $2n-2$ edges.*

Proof. Let $G_A = (V,A)$ be a 3-augmentable planar straight-line matching with $n \geq 4$ vertices. By the results of Hoffmann and Tóth [5], there is a PSLG Hamiltonian cycle H on the vertex set V that does not cross any edge in A. Since the Hamiltonian cycle H is crossing-free, none of its edges is a chord of the convex hull of vertices (otherwise, the removal of this edge would disconnect H). Hence, both (V,H) and $(V, A\cup H)$ are 3-augmentable [13]. That is, there is a 3-connected PSLG $G_C = (V,C)$ such that $A \cup H \subset C$. Lemma 3.1 completes the proof. □

4 A Collection of Simple Polygons

In this section, we prove Theorem 2. We are given a 2-regular PSLG $G_A = (V,A)$ with $n \geq 4$ vertices and n edges. If G_A is 3-augmentable, then it is contained in some 3-connected PSLG $G_C = (V,C)$, say a triangulation of G_A, which may have up to $3n-6$ edges. We will construct an augmentation $G_B = (V,B)$, $A \subseteq B \subseteq C$, with $|B| \leq 2n$ edges.

Lemma 4.1. *Let $G_A = (V,A)$ be a 2-regular graph with $n \geq 4$ vertices, and let $G_C = (V,C)$ be a 3-connected PSLG with $A \subseteq C$ such that all bounded faces are triangles. Then $G_A = (V,A)$ can be augmented to a 3-connected graph $G_B = (V,B)$ with $A \subseteq B \subseteq C$, such that $|B| \leq 2n$.*

Proof. Since G_C is 3-connected, its outer face is a simple polygon, which we denote by Q_C, and at least one vertex of V lies in the interior of Q_C. We construct a 3-connected graph G_B, $A \subseteq B \subseteq C$, incrementally. We maintain a 2-connected graph $G_i = (V_i, E_i)$ with $V_i \subseteq V$ and $E_i \subseteq C$. We also maintain a set $U_i \subseteq V$ of vertices, called *hubs*, which is the set of *all* vertices in V_i with degree 3 or higher in G_i. The hubs naturally decompose G_i into a set $\mathcal{P}_i$ of paths in G_i between hubs. We maintain the following invariants for G_i.

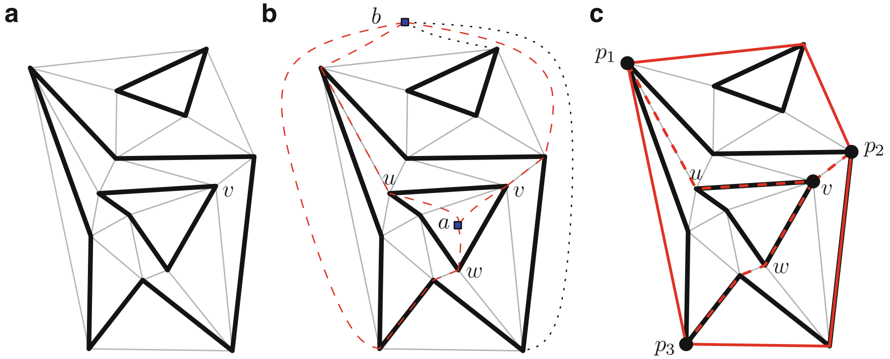


Fig. 9 (**a**) A 2-regular PSLG G_A (black) in a 3-connected triangulation G_C (*gray*). (**b**) Graph G_C^* with two auxiliary vertices, a and b, is also 3-connected. (**c**) Three independent paths from v to three boundary points p_1, p_2, and p_3

J1 $Q_C \subseteq E_i \subseteq C$.
J2 U_i is 3-linked in G_i.
J3 every bounded face of G_i is incident to at least three vertices in U_i.
J4 no edge of $C \setminus E_i$ joins a pair of vertices of any path in $\mathcal{P}_i$.

Initially, G_0 will have four vertices, and we incrementally augment it with new edges *and* vertices, until we have $U_i = V$. The vertex sets V_i, U_i, and the edge set E_i will monotonically increase during this algorithm, and we gradually add all edges of A to E_i. When our algorithm terminates and $U_i = V$, the graph G_i is a 3-connected subgraph of G_C, which contains all edges of A. Whenever we add an edge $e \in C \setminus A$ to E_i, we charge e to one of the endpoints of e so that every vertex is charged at most once. This charging scheme ensures that we add at most n edges from $C \setminus A$. Together with the n edges of G_A, we obtain a 3-connected augmentation with at most $2n$ edges.

Initialization. We construct the initial graph G_0 with $|U_0| = 4$ hubs. Consider the 3-connected PSLG G_C where Q_C is the boundary of the outer face. Let $v \in V$ be a vertex in the interior of Q_C, and let u and w be its two neighbors in the 2-regular graph G_A.

Construct an auxiliary graph $G_C^* = (V \cup \{a,b\}, C^*)$, with $C \subset C^*$, as follows. The edges of G_C^* are all edges in C, edges au, av, and aw, and edges connecting the auxiliary vertex b to all vertices of the outer face Q_C (Fig. 9b). By Lemma 2.1, G_C^* is 3-connected (albeit not necessarily planar). Hence, G_C^* contains three independent paths between a and b. Fix three independent paths of minimal total length. The minimality implies that no two nonconsecutive vertices in any of the three paths are joined by an edge of C. Replace the edges au and aw with vu and vw, respectively, to obtain three independent paths *in* C from $v \in V$ to three distinct vertices of the outer face Q_C, such that two of these paths leave v along edges of A. Denote by P_1, P_2, P_3 the three paths, with endpoints p_1, p_2, p_3 along Q_C, respectively.

Let our initial graph $G_0 = (V_0, E_0)$ consist of all edges and vertices of $Q_C \cup P_1 \cup P_2 \cup P_3$. There are exactly four vertices of degree 3, namely, $U_0 = \{v, p_1, p_2, p_3\}$, which are 3-linked in G_0. Each of the three bounded faces of G_0 is incident to 3 hubs. So G_0 satisfies invariants $\mathbf{J}_1$–$\mathbf{J}_3$. For invariant $\mathbf{J}_4$, note also that no edge in C joins nonadjacent vertices of Q_C; otherwise, G_C would not be 3-connected.

Let's estimate how many edges of G_0 are from $C \setminus A$. Orient Q_C counterclockwise, and charge every edge $e \in C \setminus A$ along Q_C to its origin. Clearly, every vertex in Q_C is charged at most once. Direct the paths P_1, P_2, and P_3 from v to p_1, p_2, and p_3, and charge each edge $e \in C \setminus A$ along the paths to its origin. Since two paths leave v along edges of A, vertex v is charged exactly once. All interior vertices of the three paths are charged at most once because the paths are independent.

Our algorithm proceeds in three phases.

Phase 1. In the first phase of our algorithm, we augment $G_i = (V_i, E_i)$ until $V_i = V$, but at the end of this phase some edges of A may still not be contained in E_i. We augment $G_i = (V_i, E_i)$ with new edges and vertices incrementally. It is enough to describe a general step of this phase.

Pick an arbitrary vertex $v \in V \setminus V_i$. We will augment G_i to include v (and possibly other vertices). Our argument is similar to the initialization. Let Q_v denote the boundary of the face of G_i that contains v. Let U_v denote the set of hubs along Q_v. We have $|U_v| \geq 3$ by $\mathbf{J}_3$, and Q_v consists of at least three paths from $\mathcal{P}_i$ between consecutive hubs along Q_v. Let G_v be the subgraph of C that contains all edges and vertices of G_C lying in the closed polygonal domain bounded by Q_v. Note that G_v is a (Steiner) triangulation of the simple polygon Q_v. This implies that G_v is 2-connected and every 2-cut of G_v consists of a pair of vertices joined by a chord of Q_v. By invariant $\mathbf{J}_4$, however, there is a hub in U_v on each side of every chord of Q_v.

Let u and w be the neighbors of v in the 2-regular graph G_A. Note that both u and w must be vertices of G_v. Construct an auxiliary graph G_v^* as follows. The vertices of G_v^* are the vertices of G_v and two auxiliary vertices, a and b. The edges of G_v^* are the edges of G_v; the edges au, av, and aw; and edges between b and every hub in U_v. We show that G_v^* is 3-connected. First note that none of the 2-cuts of G_v is a 2-cut in G_v^*, since the hubs on the two sides of a chord of Q_v are now joined to b. This implies that the vertices of G_v are 3-linked in G_v^*. Vertices a and b are each joined to three vertices of G_v, and by Lemma 2.1 there are three independent paths between any two vertices of G_v^*.

Choose three independent paths in G_v^* between a and b of minimal total length. The minimality implies that each path goes from a to a vertex along Q_v, then follows Q_v to a hub in U_v, and reaches b along a single edge from Q_v. In particular, no two nonconsecutive vertices of any of the three paths are joined by an edge of G_v^* (i.e., no shortcuts). Replace the edges au and aw with edges vu and vw, respectively, to obtain three independent paths from v to three distinct hubs along Q_v, such that two of these paths leave v along edges of A. Denote by P_1, P_2, and P_3 the initial portions of the paths between a and Q_v; and let p_1, p_2, and p_3 be their endpoints on Q_v (these endpoints are not necessarily in U_i).

We construct G_{i+1} by augmenting G_i with all vertices and edges of the paths P_1, P_2, and P_3. The new vertices of degree 3 are v and, if they were not hubs already, p_1, p_2, and p_3. In G_{i+1}, three independent paths connect v to three hubs in U_i, so $U_i \cup \{v\}$ is 3-linked in G_{i+1}. Similarly, p_1, p_2, and p_3 are each connected to three hubs in $U_i \cup \{v\}$ along three independent paths. We conclude that $U_{i+1} = U_i \cup \{u, p_1, p_2, p_3\}$ is 3-linked in G_{i+1}. We can construct $\mathcal{P}_{i+1}$ from $\mathcal{P}_i$ by adding the three new paths P_1, P_2, and P_3; and splitting the paths in $\mathcal{P}_i$ containing p_1, p_2, and p_3 into two pieces if necessary.

Paths P_1, P_2, and P_3 decompose a face of G_i into three faces, each of which is incident to at least three hubs of U_{i+1}. So invariants $\mathbf{J}_1$–$\mathbf{J}_4$ hold for G_{i+1}. It remains to charge the new edges of E_{i+1} taken from $C \setminus A$ to some new vertices in V_{i+1}. Direct the paths P_1, P_2, and P_3 from v to p_1, p_2, and p_3, respectively, and charge any new edge $e \in C \setminus A$ to its origin. Each new vertex of V_{i+1} is charged at most once: v is charged at most once because two incident new edges are contained in A; and any other new vertex is charged at most once because the paths P_1, P_2, and P_2 are independent.

Phase 2. In the second phase, we augment $G_i = (V, E_i)$ with edges of $A \setminus E_i$ successively until $A \subseteq E_i$. We can add all edges of A at no charge; we only need to check that that invariants $\mathbf{J}_1$–$\mathbf{J}_4$ are maintained. We describe a single step of the augmentation. Consider an edge $pq \in A \setminus E_i$. Let $G_{i+1} = (V, E_{i+1})$ with $E_{i+1} = E_i \cup \{pq\}$ and $U_{i+1} = U_i \cup \{p, q\}$. By $\mathbf{J}_2 - -\mathbf{J}_4$, and Lemma 2.1, U_{i+1} is 3-linked in G_{i+1}. The edge pq subdivides a bounded face of G_i into two faces of G_{i+1}. Since pq does not join two vertices of the same path in $\mathcal{P}_i$ by invariant $\mathbf{J}_4$, both new faces are incident to at least three hubs in U_{i+1} (including p and q). The paths in $\mathcal{P}_{i+1}$ are obtained from $\mathcal{P}_i$ by adding the 1-edge path pq and possibly decomposing the paths containing p and q into two. Since $\mathcal{P}_i$ satisfies $\mathbf{J}_4$, no edge in C joins two nonconsecutive vertices of any path in $\mathcal{P}_{i+1}$ either. So invariants $\mathbf{J}_1$–$\mathbf{J}_4$ hold for G_{i+1}.

Phase 3. We have a graph $G_i = (V, E_i)$ with $A \subseteq E_i \subseteq C$, where U_i is the set of vertices of degree 3 or higher. Let $W = V \setminus U_i$ be the set of vertices that have degree 2 or less than 2 in G_i. Since G_A is 2-regular, and $A \subseteq E_i$, every vertex in W has degree 2 and is incident to two edges in A. Since we charged every edge $E_i \cap (C \setminus A)$ to an incident vertex, no vertex in W has been charged so far. Apply Lemma 2.2 to augment $G_i = (V, E_i)$ to a 3-connected graph G_B with $|W|$ additional edges. Charge the new edges to the vertices in W. At the end of phase 3, every vertex is charged at most once, and we obtain a 3-connected PSLG with at most $|A| + n = 2n$ edges. □

5 Obstacles in a Container

In this section, we consider augmenting a PSLGs $G_0 = (V, E_0)$ with $n \geq 6$ vertices that consists of a set of interior-disjoint convex polygons (obstacles) in the interior of a triangular container. Since no edge is a proper chord of the convex hull, every

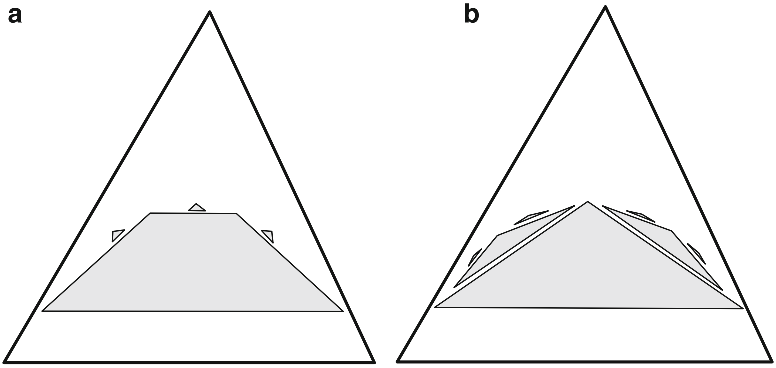


Fig. 10 (**a**) 3-Connectivity augmentation for k interior-disjoint convex obstacles in a triangular container requires $3k-1$ new edges. (**b**) For k interior-disjoint triangular obstacles in a triangular container, we need $(5k+1)/2$ new edges

such PSLG is 3-augmentable [13], and by Theorem 2 it can be augmented to a 3-connected PSLG with at most $2n$ edges. We believe, however, that significantly fewer edges are sufficient for 3-connectivity augmentation. The best lower bounds we were able to construct require fewer than $2n-2$ edges.

When there is only one convex obstacle, three edges are obviously required for connecting it to the container. However, for $k \in \mathbb{N}$ convex obstacles at least $3k-1$ edges are necessary in the worst case. Our lower bound construction is depicted in Fig. 10a. It includes one large convex obstacle, which hides one small obstacle behind each side (except the base) such that each small obstacle can "see" only three different vertices (the top vertex of the container and two adjacent vertices of the large obstacle). Thus, we need three edges for each small obstacle and only two edges for the larger obstacle, connecting its two bottom vertices to the two endpoints of the base of the container.

The large obstacle in the above construction is a convex k-gon, and so the lower bound $3k-1$ does not hold if every obstacle has at most s sides, for some fixed $3 \leq s < k$. In that case we use a similar construction, in which a big s sided, obstacle hides $s-1$ smaller obstacles behind all its sides except one, and the construction is repeated recursively. This construction corresponds to a complete tree with branching factor $s-1$, in which the smaller obstacles are the children of a larger obstacle. For a fixed value of s, we set h as the height of the complete $(s-1)$-ary tree. Thus, the number of obstacles,

$$k = \frac{(s-1)^h - 1}{s-2},$$

can be as high as we desire. The number of leaves in the tree is $(s-1)^{h-1}$. A simple manipulation of this equation shows that this number equals $k - \frac{k-1}{s-1}$. Hence, the number of internal nodes in the tree is $\frac{k-1}{s-1}$. For the 3-connectivity augmentation, each leaf obstacle needs at least s new edges and each nonleaf obstacle needs at least two new edges. The total number of edges required is at least

$$s\left(k-\frac{k-1}{s-1}\right)+2\left(\frac{k-1}{s-1}\right)=sk-\frac{s-2}{s-1}(k-1)=(n-3)-\frac{s-2}{s-1}\cdot\left(\frac{n-3}{s}-1\right),$$

which ranges from $\frac{5}{6}n-\frac{5}{2}$ to $n-O(\sqrt{n})$ for $3\le s\le\sqrt{n-3}$. Figure 10b depicts this lower bound construction for $s=3$.

6 Discussion

We have shown that a 1-regular (respectively, 2-regular) PSLG with n vertices, where no edge is a chord of the convex hull, can be augmented to a 3-connected PSLG that has at most $2n-2$ (respectively, $2n$) edges. We conjecture that our results generalize to PSLGs with maximum degree at most 2 (Conjecture 1.1).

The bound of $2n-2$ for the number of edges is the best possible in general, but it may be improved if few vertices lie on the convex hull, and the components of the input graph are interior-disjoint convex obstacles, possibly with a container. It remains an open problem to derive tight extremal bounds for 3-connectivity augmentation for (a) 1-regular PSLGs with n vertices, h of which lie on the convex hull; and (b) 2-regular PSLGs formed by $\frac{n}{s}$ interior-disjoint convex polygons, each with s vertices for $s\ge 3$.

The 3-connectivity augmentation problem (finding the *minimum* number of new edges for a given PSLG) is known to be NP-hard [12]. However, the hardness proof does not apply to 1- or 2-regular polygons. It is an open problem whether the connectivity augmentation remains NP-hard restricted to these cases.

We have compared the number of edges in the resulting 3-connected PSLGs with the benchmark $2n-2$, which is the best possible bound for 0-, 1-, and 2-regular PSLGs. More generally, for a 3-augmentable PSLG $G_0=(V,E_0)$ with $n\ge 4$ vertices, let $f(G_0)=|E_1|$ be the minimum number of edges in a 3-connected augmentation (V,E_1) of the *empty* PSLG $(V,\emptyset)$; and let $g(G_0)=|E_2|$ be the minimum number of edges in a 3-connected augmentation (V,E_2), $E_0\subseteq E_2$, of the PSLG G_0. It is clear that $f(G_0)\le g(G_0)$. With this notation, we can characterize the PSLGs G_0, where *all* edges in E_0 are "useful" for 3-connectivity: These are the PSLGs for which $f(G_0)=g(G_0)$ is possible. In general, it would be interesting to study the behavior of the difference $g(G_0)-f(G_0)$.

Finally, we note here that augmenting a 1-regular PSLG to a 2-regular PSLG has been considered by Aichholzer et al. [1] and Ishaque et al. [8]. The problem is not always feasible if the input graph has an odd number of edges, but neither combinatorial characterizations nor hardness results are known for the corresponding decision problem.

Acknowledgements We are grateful to the anonymous referee who pointed out several errors and omissions in an earlier version of the proof of Lemma 3.1.

This material is based upon work supported by the National Science Foundation under Grant No. 0830734. Research by C. D. Tóth was also supported by NSERC grant RGPIN 35586. Preliminary results have been presented at the *26th European Workshop on Computational Geometry (2010, Dortmund)* and at the *20th Annual Fall Workshop on Computational Geometry (2010, Stony Brook, NY).*

References

1. O. Aichholzer, S. Bereg, A. Dumitrescu, A. García, C. Huemer, F. Hurtado, M. Kano, A. Márquez, D. Rappaport, S. Smorodinsky, D. L. Souvaine, J. Urrutia, D.R. Wood, Compatible geometric matchings. Comput. Geom.: Theor. Appl. **42**, 617–626 (2009)
2. M. Al-Jubeh, M. Ishaque, K. Rédei, D.L. Souvaine, C.D. Tóth, Augmenting the edge connectivity of planar straight line graphs to three. Algorithmica
3. K.P. Eswaran, R.E. Tarjan, Augmentation problems. SIAM J. Comput. **5**, 653–665 (1976)
4. A. García, F. Hurtado, C. Huemer, J. Tejel, P. Valtr, On triconnected and cubic plane graphs on given point sets. Comput. Geom.: Theor. Appl. **42**, 913–922 (2009)
5. M. Hoffmann, C.D. Tóth, Segment endpoint visibility graphs are Hamiltonian. Comput. Geom.: Theor. Appl. **26**, 47–68 (2003)
6. T.-S. Hsu, V. Ramachandran, On finding a minimum augmentation to biconnect a graph. SIAM J. Comput. **22**(5), 889–912 (1993)
7. T.-S. Hsu, Simpler and faster biconnectivity augmentation. J. Algorithms **45**(1), 55–71 (2002)
8. M. Ishaque, D.L. Souvaine, C.D. Tth, Disjoint compatible geometric matchings, Discrete and Computational Geometry, in print. DOI 10.1007/s00454-012-9466-9 (2012)
9. B. Jackson, T. Jordán, Independence free graphs and vertex connectivity augmentation. J. Comb. Theor. Ser. B **94**, 31–77 (2005)
10. G. Kant, H.L. Bodlaender, Planar graph augmentation problems, in *Proceedings of the 2nd Workshop on Algorithms and Data Structures*. Lecture Notes in Computer Science, vol. 519 (Springer-Verlag, Berlin, 1991), pp. 286–298
11. J. Plesník, Minimum block containing a given graph. Arch. Math. **21**, 668–672 (1976)
12. I. Rutter, A. Wolff, Augmenting the connectivity of planar and geometric graphs. in *Proceedings of the International Conference on Topological and Geometric Graph Theory*, Paris. Electron. Notes Discr. Math. **31**, 53–56 (2008)
13. C.D. Tóth, P. Valtr, Augmenting the edge connectivity of planar straight line graphs to three, in *13th Spanish Meeting on Computational Geometry*, Zaragoza, Spain, 2009

Topological Hypergraphs

Sarit Buzaglo, Rom Pinchasi, and Günter Rote

Abstract Let P be a set of n points in the plane. A *topological hypergraph* G, on the set of points of P, is a collection of simple closed curves in the plane that avoid the points of P. Each of these curves is called an *edge* of G, and the points of P are called the *vertices* of G. We provide bounds on the number of edges of topological hypergraphs in terms of the number of their vertices under various restrictions assuming the set of edges is a family of pseudo-circles.

1 Introduction

A *topological graph* is a graph drawn in the plane with its vertices drawn as points and its edges drawn as Jordan arcs connecting corresponding points. In this chapter we make an attempt to generalize the notion of a topological graph and consider topological *hypergraphs*.

Definition 1.1. Let C be a simple closed Jordan curve in the plane. By Jordan's theorem, C divides the plane into two regions, only one of which is bounded. We call the bounded region the *disc* bounded by C and we denote this region by $\mathrm{disc}(C)$. Any point x inside $\mathrm{disc}(C)$ is said to be *surrounded* by C, and C is said to be *surrounding* x.

We are now ready to define a topological hypergraph. A *topological hypergraph* G is a pair $(P,\mathcal{C})$, where P is a finite set of points in the plane, and $\mathcal{C}$ is a family of

S. Buzaglo • R. Pinchasi (✉)
Technion—Israel Institute of Technology, Mathematics Department, Haifa 32000, Israel
e-mail: sarahb@tx.technion.ac.il; room@math.technion.ac.il

G. Rote
Institut für Informatik, Freie Universität Berlin, Takustr. 9, 14195 Berlin, Germany
e-mail: rote@inf.fu-berlin.de

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_6,

simple closed curves in the plane that avoid the points of P. For each curve $C \in \mathcal{C}$, we denote by P_C the set of all points of P surrounded by C. C is called an *edge* of G and the set P_C is the set of *vertices* of the edge C. Two edges of G, C_1 and C_2, are called *parallel* if $P_{C_1} = P_{C_2}$. We usually assume that any two curves from $\mathcal{C}$ intersect in a finite number of points and that in each such intersection point the two curves properly cross. G is called k-uniform if, for every $C \in \mathcal{C}$, $|P_C| = k$.

In what follows and throughout the rest of the chapter, we will always assume that the topological hypergraph in question *does not contain parallel edges*.

A family of simple closed curves in the plane is a family of *pseudo-circles* if every two curves in the family are either disjoint or properly cross at precisely two points. Any collection of circles in the plane is an example of such a family.

The following simple lemma is a crucial observation.

Lemma 1. *Suppose C_1 and C_2 are two pseudo-circles in the plane and x_1, y_1, x_2, y_2 are four distinct points satisfying the following condition: x_1 and y_1 are surrounded by C_1 but not by C_2, and x_2 and y_2 are surrounded by C_2 but not by C_1. Let e_1 be a Jordan arc, entirely contained in disc(C_1), connecting x_1 and y_1, and let e_2 be a Jordan arc, entirely contained in disc(C_2), connecting x_2 and y_2. Then e_1 and e_2 cross an even number of times.*

Proof. If C_1 and C_2 do not intersect, then $\mathrm{disc}(C_1)$ and $\mathrm{disc}(C_2)$ must be disjoint. Indeed, otherwise one must contain the other; say $\mathrm{disc}(C_1)$ contains $\mathrm{disc}(C_2)$. This is a contradiction to the assumption that x_2 and y_2 are two points surrounded by C_2 but not by C_1. However, if $\mathrm{disc}(C_1)$ is disjoint from $\mathrm{disc}(C_2)$, then clearly e_1 and e_2 cannot intersect.

Therefore, assume that C_1 and C_2 intersect in two points u and v. Let S_1 be the subarc of C_1 delimited by u and v and contained in $\mathrm{disc}(C_2)$. Denote by S_2 the subarc of C_2 delimited by u and v and contained in $\mathrm{disc}(C_1)$. Observe that S_1 together with S_2 form a simple closed curve that surrounds exactly the points in $Z = \mathrm{disc}(C_1) \cap \mathrm{disc}(C_2)$. Clearly, e_1 and e_2 may intersect only at points of Z (see Fig. 1). Since both x_1 and y_1 do not belong to Z, as they are not surrounded by C_2, the intersection of e_1 with Z consists of a finite number of arcs each of which has both endpoints lying on S_2. It is enough to show that each of these arcs crosses e_2 an even number of times. Let g be such an arc whose endpoints on S_2 are k and ℓ. Let S'_2 be the subarc of S_2 delimited by k and ℓ. g and S'_2 form a simple closed curve contained in Z. Consequently, both x_2 and y_2 are not surrounded by this closed curve. It follows that e_2, the Jordan arc connecting x_2 and y_2, intersects $g \cup S'_2$ an even number of times. However, e_2 does not intersect S'_2, as it is a subarc of C_2. Therefore, e_2 crosses g an even number of times. ∎

The next theorem is a simple consequence of Lemma 1.

Theorem 2. *Let $G = (P, \mathcal{C})$ be a topological 2-uniform hypergraph on n vertices. If $\mathcal{C}$ is a family of pseudo-circles, then G, viewed as an abstract 2-graph, is a planar graph. In particular, there are at most $3n-6$ curves in $\mathcal{C}$.*

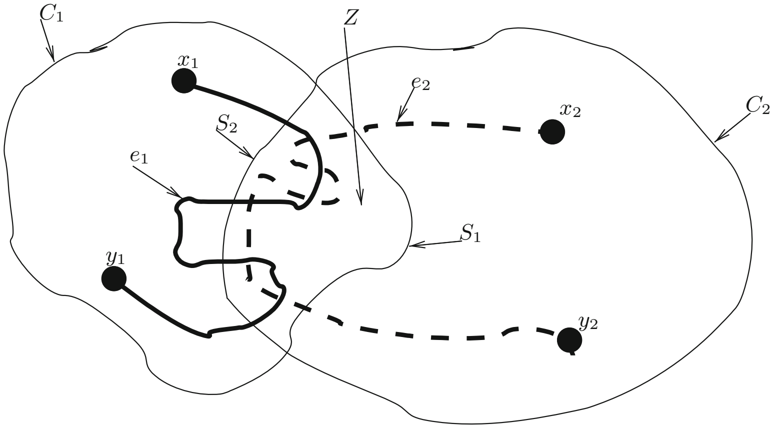


Fig. 1 Lemma 1

Proof. We draw the graph G as a topological graph in such a way that every pair of edges that are not incident to the same vertex cross an even number of times. Then it is a consequence of Hanani and Tutte's theorem [2, 9] that G is planar. The drawing rule for G is as follows: For any curve $C \in \mathcal{C}$, let x and y be the two points of P surrounded by C. We draw a Jordan arc connecting x to y so that the arc is contained in disc(C).

It remains to show that the above drawing rule provides the desired topological graph. Let e and f be two edges in our drawing that do not share a common vertex. Let x and y be the vertices of e, and let $C_e \in \mathcal{C}$ be the curve in $\mathcal{C}$ that surrounds both x and y. Let a and b be the vertices of f, and let $C_f \in \mathcal{C}$ be the curve in $\mathcal{C}$ that surrounds both a and b. Since C_e surrounds x and y and no other vertex, then in particular it does not surround a or b. Similarly, C_f surrounds a and b but none of x and y. Therefore, by Lemma 1, e and f cross an even number of times. ∎

We will now bound the number of edges in a topological hypergraph with more general settings. We use the following lemma [1] (see also [5]).

Lemma 3. *Let $A_1, \ldots, A_n$ be connected sets in the plane, each of which is also simply connected with boundaries that cross each other a finite number of times. If, for every $1 \le i < j \le n$, $A_i \cap A_j$ is connected, then $A_1 \cap \ldots \cap A_n$ is either connected or empty.*

As a consequence of Lemma 3, we obtain the following corollary:

Corollary 4. *If $C_1, \ldots, C_t$ is any collection of pseudo-circles, then disc$(C_1) \cap \ldots \cap$ disc(C_t) is either a connected set or the empty set.*

Indeed, this is because disc(C_i) is a simply connected set for every C_i, and disc$(C_i) \cap$ disc(C_j) is either empty or connected for every C_i and C_j.

The next theorem provides a linear bound in terms of $|P|$ on the number of edges of a 3-uniform topological hypergraph $G = (P,\mathcal{C})$, in the case where $\mathcal{C}$ is a family of pseudo-circles.

Theorem 5. *Let $G = (P,\mathcal{C})$ be a 3-uniform topological graph on n vertices such that $\mathcal{C}$ is a family of pseudo-circles. Then $|\mathcal{C}| = O(n)$.*

Proof. We draw a topological graph H whose vertices are the points of P and whose edges are all pairs of points that are surrounded by some curve in $\mathcal{C}$. Let x and y be two points of P surrounded by a curve in $\mathcal{C}$, and let $C_1,\ldots,C_l$ be all the curves in $\mathcal{C}$ that surround both x and y. By Corollary 4, $Z = \mathrm{disc}(C_1)\cap\ldots\cap\mathrm{disc}(C_l)$ is a connected set. We draw the edge in H between x and y as a Jordan curve inside the connected region Z.

For every edge e in H, let $d(e)$ denote the number of curves in $\mathcal{C}$ that surround both endpoints of e (and therefore also the entire edge e).

Claim 1. Let e and f be two edges of H with no common vertex. If e and f cross an odd number of times, then either $d(e) \leq 2$, or $d(f) \leq 2$, or both.

Proof. Assume to the contrary that both $d(e)$ and $d(f)$ are greater than or equal to 3. Since $d(e) \geq 3$, there are at least three different curves from $\mathcal{C}$ that surround both endpoints of e. Therefore, there is a curve $C_e \in \mathcal{C}$ that surrounds both endpoints of e but no endpoint of f. Similarly, there is a curve $C_f \in \mathcal{C}$ that surrounds both endpoints of f but no endpoint of e. By Lemma 1, e and f cross an even number of times, a contradiction. ∎

Let E' be the set of all edges e of H such that $d(e) \leq 2$. By Claim 1, the set of edges in $E \setminus E'$ forms a planar graph, as every two edges in $E \setminus E'$ that do not share a common vertex cross an even number of times. In particular, the cardinality of $E \setminus E'$ is linear in n. We will now show that the cardinality of E' is also linear in n.

In fact, we will prove even a slightly stronger statement that will be helpful for the proof of Theorem 5.

Claim 7. Let V be any subset of the vertices of H and let $E'(V)$ be those edges in E' both of whose vertices are in V. Then $|E'(V)| \leq c|V|$, where $c > 0$ is some absolute constant.

Proof. Let $0 < q < 1$ be a positive number to be determined later. Pick every vertex of V with probability q and consider only those edges of $E'(V)$ both of whose vertices are picked. We thus obtain a random subgraph of H, which we denote by $\tilde{H}$, on a set of vertices $\tilde{V}$. Let e be an edge in $\tilde{H}$, and let x and y be its two vertices. We know that $d(e) \leq 2$. Therefore, there exist at most two vertices $z \in V$ such that x,y, and z are surrounded by some curve $C \in \mathcal{C}$. We say that e is *good* if there is no such vertex z in $\tilde{V}$.

Observe that every two good edges with no common vertex cross an even number of times. Indeed, if e and f are two good edges with no common vertex, then there is a curve $C_e \in \mathcal{C}$ that surrounds both vertices of e but no vertex of f (or else e would not be good). And similarly, there is a curve $C_f \in \mathcal{C}$ that surrounds both vertices of f but no vertex of e. Hence, by Lemma 1, e and f cross an even number of times. It follows now from the theorem of Hanani and Tutte [2, 9] that the good edges in $\tilde{H}$ constitute a planar graph. Therefore, the number of good edges in $\tilde{G}$ is at most $3|\tilde{V}|$. This inequality is true also for the expected values of the number of good edges and $|\tilde{V}|$. Clearly, $Ex(|\tilde{V}|) = q|V|$. As for the expected number of good edges, observe that any edge $e \in E(V)$ is a good edge in $\tilde{H}$ with probability of at least $q^2(1-q)^2$. Indeed, this is the probability in case there are precisely two edges in G that include the vertices of e. If there is only one such edge in G, this probability is be higher, namely, $q^2(1-q)$.

It follows now that $q^2(1-q)^2|E(V)| \leq 3q|V|$. Taking $q = \frac{1}{3}$, we obtain $|E(V)| \leq 21|V|$. ∎

Considering the graph H again, we deduce that there is an absolute constant c' such that the number of edges of H induced by any subset V of vertices of H is at most $c'|V|$. Indeed, this is true for the subset E' of edges of H, by Claim 7, and also to the subset of edges $E \setminus E'$, as they constitute a planar graph.

We now show that the number of triangles (that is, cycles of length 3) in the graph H is at most linear in the number of vertices of H. This will be enough to prove Theorem 5 because every pseudo-circle $C \in \mathcal{C}$ corresponds to precisely one triangle in the graph G (the triangle whose vertices are surrounded by C). We prove a more general lemma on abstract graphs:

Lemma 8. *Let H be a graph on m vertices. Assume that there is an absolute constant c' such that the number of edges of H induced by any subset V of the vertices of H is at most $c'|V|$. Then for every $\ell \geq 2$, the number of copies of K_ℓ (the complete graph on ℓ vertices) in H is at most $c_\ell m$, where $c_\ell = \frac{(2c')^{\ell-1}}{\ell!}$.*

Proof. We prove the lemma by induction on ℓ. For $\ell = 2$, there is nothing to prove, as this is the assumption in the lemma. Assume the lemma is true for ℓ. Let $v_1, \ldots, v_m$ denote the vertices of H and let E denote the set of edges of H. We know that $|E| \leq c'm$. For every $1 \leq i \leq m$, let d_i denote the degree of v_i in H. Fix i and consider the number of copies of $K_{\ell+1}$ in H that include the vertex v_i. Each such copy of $K_{\ell+1}$ corresponds to a unique copy of K_ℓ among the neighbors of v_i. Therefore, it follows from the induction hypothesis, applied to the subgraph H_{v_i} of H induced by the neighbors of v_i, that the number of copies of K_ℓ in H_{v_i} is at most $c_\ell d_i$. It follows now that that v_i is incident to at most $c_\ell d_i$ copies of $K_{\ell+1}$ in H. Summing over all i between 1 and m, every copy of $K_{\ell+1}$ is counted exactly $\ell+1$ times. Therefore, the number of copies of $K_{\ell+1}$ in H is at most $\frac{1}{\ell+1}\sum_{i=1}^{m} c_\ell d_i = \frac{1}{\ell+1} 2c_\ell |E| \leq \frac{1}{\ell+1} 2c_\ell c' m = c_{\ell+1} m$. ∎

In particular, the number of triangles in our topological graph H is at most cn, where c is some absolute constant independent of n. This concludes the proof of Theorem 5.

If $G = (P, \mathcal{C})$ is a topological hypergraph and $\mathcal{C}$ is a family of pseudo-circles, but we do not assume that G is uniform, then a linear bound on the number of edges of

G is no longer valid. Nevertheless, it is possible to obtain a cubic (that is also tight) bound on the number of edges in G in this case, as we shall now see.

Recall that for a family $\mathcal{F}$ of sets, the *VC-dimension* [10] of $\mathcal{F}$ is the largest cardinality of a set S that is *shattered* by $\mathcal{F}$, that is, the largest cardinality of a set S such that for any subset B of S there exists $F \in \mathcal{F}$ with $B = F \cap S$.

Perhaps one of the most fundamental results on VC-dimension is the Perles–Sauer–Shelah theorem [7, 8], which says that a family $\mathcal{F}$ of subsets of $\{1,\ldots,n\}$ that has VC-dimension d consists of at most $\binom{n}{0} + \cdots + \binom{n}{d} = O(n^d)$ members.

Theorem 9. *Let $G = (P,\mathcal{C})$ be a topological hypergraph on n vertices such that $\mathcal{C}$ is a family of pseudo-circles. Then the family $\mathcal{F} = \{P_C \mid C \in \mathcal{C}\}$ has VC-dimension at most* 3. *In particular, by the Perles–Sauer–Shelah theorem, $|\mathcal{C}| = O(n^3)$.*

Proof. We show that $\mathcal{F}$ cannot shatter any set of 4 points. Assume to the contrary that it does, and let $\{p_1, p_2, p_3, p_4\} \subset P$ be a set of four points shattered by $\mathcal{F}$. For every $1 \le i < j \le 4$, consider the family $\mathcal{C}_{ij}$ of all the curves in $\mathcal{C}$ that surround both p_i and p_j. By Corollary 4, the set $R_{ij} = \cap_{C \in \mathcal{C}_{ij}} \mathrm{disc}(C)$ is a connected set. We draw an edge (Jordan arc) e_{ij} between p_i and p_j inside the region R_{ij}. We thus get a drawing of K_4 in the plane. We claim that in this drawing every two edges that do not share a common vertex cross an even number of times. Indeed, consider the edge e_{ij} between p_i and p_j and the edge e_{kl} between p_k and p_l (we assume that $\{i, j, k, l\} = \{1,2,3,4\}$). As $\{p_1, p_2, p_3, p_4\}$ is shattered by $\mathcal{F}$, there is a curve $C_{ij} \in \mathcal{C}_{ij}$ that surrounds both p_i and p_j but neither p_k nor p_l. Similarly, there is a curve $C_{kl} \in \mathcal{C}_{kl}$ that surrounds both p_k and p_l but neither p_i nor p_j. By our construction, $e_{ij} \subset \mathrm{disc}(C_{ij})$ and $e_{kl} \subset \mathrm{disc}(C_{kl})$. It follows now from Lemma 1 that e_{ij} and e_{kl} cross an even number of times.

It is easy to observe that by small modifications of the drawing of the edges in a small neighborhood around each point p_i, we may obtain a new drawing in which also every two edges that share a common vertex cross an even number of times (apart from meeting at the same vertex).

For a closed curve S in the plane, not necessarily simple, and a point x not on S, we say that x is *surrounded* by S if the index of the curve S with respect to the point x is odd. This is equivalent to that any ray (or just a Jordan curve that goes to infinity) emanating from x meets S (as a curve, not as a set) an odd number of times. Observe that this definition generalizes the notion of a point surrounded by a simple closed curve.

Lemma 10. *In any drawing of K_4 in the plane in which every two edges cross an even number of times, there is a vertex v of K_4 that is surrounded by the closed curve composed from the three edges of K_4 not incident to v.*

Proof. Assume not and once again let p_1, p_2, p_3, and p_4 denote the vertices of K_4 drawn in the plane. For any three vertices p_i, p_j, and p_k, denote by S_{ijk} the closed curve composed from the edges p_ip_j, p_jp_k, and p_kp_i. Consider a fixed vertex p_i. Any small enough circular disc D, centered at p_i, is trisected by the edges going from p_i to the three other vertices. Fix such a disc D. We claim the section of D

bounded by the edges p_ip_j and p_ip_k consists of points that are all surrounded by S_{ijk}. Indeed, otherwise the portion of the edge p_ip_l inside D is surrounded by S_{ijk}. As the edge p_ip_l crosses each of the edges p_ip_j, p_ip_k, and p_jp_k an even number of times, it follows that p_l is also surrounded by S_{ijk}, a contradiction.

Consider now a point x far away, not surrounded by any of the curves S_{ijk}. Draw an arc g connecting x to p_1. Once again let D be a small circular disc centered at p_1. D is trisected by the edges p_1p_i for $i = 2,3,4$. Without loss of generality, assume that $g \cap D$ lies in the portion of D delimited by p_1p_2 and p_1p_3. Hence, as x is not surrounded by S_{123}, it follows that g crosses S_{123} an odd number of times. For the same reasons, it follows that g crosses both S_{124} and S_{134} an *even* number of times because $g \cap D$ is not surrounded by S_{124} nor by S_{134}.

Denote by t_{ij} the number of crossings between g and the edge p_ip_j. Therefore, $t_{12}+t_{13}+t_{23}$ is odd and both $t_{12}+t_{14}+t_{24}$ and $t_{13}+t_{14}+t_{34}$ are even. Summing them all up, we conclude that $t_{23}+t_{34}+t_{24}$ is odd. Therefore, g crosses S_{234} an odd number of times. Now since x is not surrounded by S_{234} and g connects x to p_1, it follows that p_1 is surrounded by S_{234}, a contradiction. ∎

In view of Lemma 10, assume without loss of generality that p_4 is surrounded by the closed curve composed from the edges p_1p_2, p_2p_3, and p_3p_1. As $\mathcal{F}$ shatters $\{p_1,p_2,p_3,p_4\}$, let $C \in \mathcal{C}$ be a curve surrounding p_1,p_2, and p_3 but not p_4. Therefore, each of the edges p_1p_2, p_2p_3, and p_3p_1 is contained in $\mathrm{disc}(C)$. This is a contradiction because now it is not possible that p_4 is surrounded by the closed curve composed from the edges p_1p_2, p_2p_3, and p_3p_1, as it is not surrounded by C. This completes the proof of Theorem 9.

For a set A and an integer $r \geq 0$, we denote by $\binom{[A]}{\leq r}$ the family of all subsets of A of cardinality at most r. We will need the following small variant of the original Perles–Sauer–Shelah theorem [7].

Theorem 11. *Let $\mathcal{F} = \{A_1,\ldots,A_m\}$ be a family of distinct subsets of $\{1,2,\ldots,n\}$ and assume that $\mathcal{F}$ has VC-dimension less than or equal to d. Then*

$$m \leq |\bigcup_{i=1}^{m} \binom{[A_i]}{\leq d}|.$$

Proof. We prove the theorem by induction on d and n. The theorem is clearly true for $d=0$ and any n. For $d>0$, assume that $\mathcal{F}$ has VC-dimension at most d and define $\mathcal{F}_1 = \{A \setminus \{1\} \mid A \in \mathcal{F}\}$ and $\mathcal{T} = \{A \setminus \{1\} \mid 1 \in A \text{ and } A \setminus \{1\} \in \mathcal{F}\}$. It is easy to see that $|\mathcal{F}| = |\mathcal{F}_1| + |\mathcal{T}|$. Let s denote the number of sets in $\mathcal{F}_1$ and let t denote the number of sets in $\mathcal{T}$. We rewrite $\mathcal{F}_1 = \{F_1,\ldots,F_s\}$ and $\mathcal{T} = \{T_1,\ldots,T_t\}$. As the VC-dimension of $\mathcal{F}_1$ is at most d, we use the induction hypothesis on n to deduce that the family $\bigcup_{i=1}^{s} \binom{[F_i]}{\leq d}$ contains at least s sets.

Observe that the VC-dimension of $\mathcal{T}$ is at most $d-1$, for otherwise $\mathcal{F}$ has VC-dimension at least $d+1$. Therefore, by the induction hypothesis, there are at least t sets in $\bigcup_{i=1}^{t} \binom{[T_i]}{\leq d-1}$. To each of these sets add the element 1 to obtain t sets in

$\bigcup_{i=1}^{m}\binom{[A_i]}{\leq d}$ each containing the element 1 and thus each is different than any set in $\bigcup_{i=1}^{s}\binom{[F_i]}{\leq d}$. The result follows from the fact that $m = s + t$. ∎

Corollary 12. *Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ be a family of distinct subsets of $\{1, 2, \ldots, n\}$ with VC-dimension less than or equal to d. Then it is possible to assign to each set A_i a subset of it of size at most d such that no two sets A_i and A_j are assigned to the same set.*

Proof. Consider the bipartite graph in which one set of vertices corresponds to the sets $A_1, \ldots, A_m$ and the other set of vertices corresponds to the elements in $\bigcup_{i=1}^{m}\binom{[A_i]}{\leq d}$. In this bipartite graph connect each set A_i to all its subsets of size at most d. By Hall's theorem [3,4] and by Theorem 11, the desired matching exists. ∎

Theorems 2 and 5 give a linear bound to the number of edges of a 2-uniform and 3-uniform, respectively, topological hypergraphs where the set of edges is a family of pseudo-circles. The next theorem provides a bound to the number of edges in a topological hypergraph $G = (P, \mathcal{C})$ on n vertices, where $\mathcal{C}$ is a family of pseudo-circles and that $|P_C| \leq k$ for every $C \in \mathcal{C}$ and a fixed $k > 3$.

Theorem 13. *Let $G = (P, \mathcal{C})$ be a topological hypergraph on n vertices. Assume that $\mathcal{C}$ is a family of pseudo-circles and that $|P_C| \leq k$ for every $C \in \mathcal{C}$ and a fixed integer k. Then $|\mathcal{C}| = O(k^2 n)$.*

Proof. Consider the family $\mathcal{F} = \{P_C \mid C \in \mathcal{C}\}$. By Theorem 9, $\mathcal{F}$ has VC-dimension less than or equal to 3. We use Corollary 12 and assign to each member P_C of $\mathcal{F}$ a subset of it of size at most 3 that we denote by B_C.

We fix a small number $0 < q < 1/2$ to be determined later and pick every point of P independently with probability q. We call a curve $C \in \mathcal{C}$ *good* if every point in B_C was picked and no point in $P_C \setminus B_C$ was picked. Observe that a curve C in $\mathcal{C}$ is good with probability $q^{|B_C|}(1-q)^{|P_C|-|B_C|} \geq q^{|B_C|}(1-q)^{k-|B_C|} \geq q^3(1-q)^{k-3}$, where the last inequality is because $q < 1/2$ and $|B_C| \leq 3$.

By Theorems 2 and 5, the number of edges in a 2-uniform, or a 3-uniform topological hypergraph on n vertices whose set of edges is a family of pseudo-circles is $O(n)$. Therefore, the number of good curves in $\mathcal{C}$ is at most some absolute constant c times the number of points of P that were picked. Taking the expectation, we see that $|\mathcal{C}|q^3(1-q)^{k-3} \leq cqn$. Taking $q = 1/k$, we get the desired result, namely, $|\mathcal{C}| \leq c'k^2 n$ for some absolute constant $c' > 0$. ∎

Remark 1. It is not hard to show that the bound in Theorem 13 can indeed be attained even by a family $\mathcal{C}$ of circles in the plane.

Definition 1.2. We denote by $f_t(n)$ the maximum number of edges in a topological graph on n vertices with no t vertex-disjoint edges every two of which cross an odd number of times.

We will need the following estimate on $f_t(n)$ from [6] (see Theorem 3 in [6]):

Theorem 14 ([6]). $f_t(n) = O(n \log^{4t-8} n)$.

In fact, in the sequel, we will only use the fact that $f_t(n) = o(n^2)$.

The next theorem provides even better bounds on the size of $\{P_C \mid C \in \mathcal{C}\}$, assuming that the intersections of the sets P_C with each other are "small."

Theorem 15. *Let $r, k_1, k_2 > 0$ be fixed integers. Let $G = (P, \mathcal{C})$ be a topological hypergraph on n vertices and assume that $\mathcal{C}$ is a family of pseudo-circles. Assume further that $k_1 \le |P_C| \le k_2$ for every $C \in \mathcal{C}$. If, for every r curves $C_1, \dots, C_r \in \mathcal{C}$, we have $|P_{C_1} \cap \ldots \cap P_{C_r}| < r$, then $\sum_{C \in \mathcal{C}} |P_C| = O(\frac{k_2^2}{k_1^2} n)$, where the constant of proportionality depends only on r.*

Proof. We define a topological graph H in the following way: The vertices of H are the points of P. For any pair of vertices x and y that are surrounded by some curve $C \in \mathcal{C}$, consider all curves in $\mathcal{C}$ that surround x and y. Let those curves be $C_1, \dots, C_l$ and define $Z = \text{disc}(C_1) \cap \ldots \cap \text{disc}(C_l)$. By Corollary 4, Z is a connected set. We draw the edge in H between x and y as a Jordan curve entirely included in Z.

We call an edge of H *bad* if its endpoints, and hence the edge as well, are surrounded by at least $(r-1)2^{r-2} + 1$ different curves in $\mathcal{C}$; otherwise, it is called *good*.

Claim 6. There are no r^2 bad edges in H every two of which cross an odd number of times and no two of which share a common vertex.

Proof. Assume to the contrary that E is a collection of r^2 bad edges every two of which cross an odd number of times and no two of which share a common vertex.

Let $e, e_1, \dots, e_{r-2}$ be any $r-1$ different edges in E. We claim that among all curves in $\mathcal{C}$ surrounding e, there is at least one curve that does not surround any of the two vertices of some edge among $e_1, \dots, e_{r-2}$. Indeed, otherwise any curve in $\mathcal{C}$ that surrounds e must surround at least one vertex of each of $e_1, \dots, e_{r-2}$. Since e is a bad edge, there are at least $(r-1)2^{r-2} + 1$ curves in $\mathcal{C}$ that surround e. Therefore, by the pigeon-hole principle, there must be at least r curves surrounding e as well as the same set of $r-2$ vertices (each of which is the endpoint of some edge among $e_1, \dots, e_{r-2}$). This is a contradiction to the assumption that no r curves in $\mathcal{C}$ surround the same set of r vertices.

It follows that apart from at most $r-3$ edges in E, for every edge f in E, different from e, there is a curve $C \in \mathcal{C}$ that surrounds e but not any of the two vertices of f. This way, because $E > (r-1)(r-2)$, we can find $r-1$ edges $e_1, \dots, e_{r-1}$ with the property that for every $1 \le i < j \le r-1$ there is a curve in $\mathcal{C}$ that surrounds e_i but not any of the two vertices of e_j.

Using once again the argument above, there is a curve $C \in C$ and there is an index j between 1 and $r-2$ such that C surrounds e_{r-1}, but C does not surround any of the vertices of e_j. Since $j < r-1$, there is a curve $C' \in \mathcal{C}$ that surrounds e_j but that does not surround any of the vertices of e_{r-1}. By Lemma 1, this implies that e_j and e_{r-1} cross an even number of times, contradicting our assumption. ∎

It follows from Claim 6 and Theorem 14 that there exists k_0 that depends only on r such that if $k_1 > k_0$, then for every $C \in \mathcal{C}$, C surrounds at least $k_1^2/4$ good edges.

Clearly, we may assume that $k_1 > k_0$ as by Theorem 13, the collection $\mathcal{C}'$ of all curves $C \in \mathcal{C}$ with $|P_C| \le k_0$ consists of $O(k_0^2 n)$ curves and therefore, $\sum_{C \in \mathcal{C}'} |C| = O(n)$.

Fix a probability $0 < q < 1$, to be determined later, and pick every point in P with probability q. Let H^* be the random topological graph whose vertices are those points that were picked and whose edges are those good edges of H both of whose vertices were picked. Call an edge e of H^* *nice* if there is a curve $C \in \mathcal{C}$ that surrounds no other vertex of H^* but the endpoints of e.

We claim that the subgraph of H^*, which consists only of the nice edges, is planar. Indeed, let e and f be two nice edges that do not share a common vertex. Because e is nice, there is a curve $C_e \in \mathcal{C}$ that surrounds e but does not surround any of the vertices of f. Similarly, there is a curve $C_f \in \mathcal{C}$ that surrounds f but does not surround any of the vertices of e. By Lemma 1, e and f cross an even number of times. Therefore, by the Hanani–Tutte theorem, H^* is planar. It follows that $|E^*|$ is less than or equal to $3|V^*| - 6$, where E^* is the set of all nice edges and V^* is the set of vertices in H^*. This is true also when considering the expected values of $|E^*|$ and $|V^*|$.

We have $Ex(|V^*|) = qn$. We claim that $Ex(|E^*|) \ge |W| q^2 (1-q)^{k_2 - 2}$, where W is the set of all good edges. This is because each edge in W is nice with probability of at least $q^2(1-q)^{k_2-2}$. Indeed, suppose $e \in W$, and let C_e be any curve in $\mathcal{C}$ that surrounds e. e becomes nice if its two vertices are picked to V^* (which happens with probability q^2) and the other points surrounded by C_e are not picked. The latter happens with probability of at least $(1-q)^{k_2-2}$ and independently of the event in which the two vertices of e are picked.

Therefore, $|W| q^2 (1-q)^{k_2-2} < 3qn$. Taking $q = \frac{1}{k_2 - 1}$, we obtain $|W| < 9(k_2 - 1)n$. On the other hand, each good edge is surrounded by at most $(r-1)2^{r-1}$ curves in $\mathcal{C}$, and each curve in $\mathcal{C}$ surrounds at least $k_1^2/4$ good edges. Therefore, $|W| \ge \frac{k_1^2}{4(r-1)2^{r-1}} |\mathcal{C}|$. Hence, $|\mathcal{C}| < \frac{36(r-1)2^{r-1} k_2}{k_1^2} n$. The theorem now follows as $\sum_{C \in \mathcal{C}} |P_C| \le \sum_{C \in \mathcal{C}} k_2 \le 36(r-1)2^{r-1} \frac{k_2^2}{k_1^2} n$. ∎

Corollary 17. *Let $r > 0$ be a fixed integer. Let $G = (P, \mathcal{C})$ be a topological hypergraph on n vertices, and assume that $\mathcal{C}$ is a family of pseudo-circles. If, for every r curves $C_1, \ldots, C_r \in \mathcal{C}$, we have $|P_{C_1} \cap \ldots \cap P_{C_r}| < r$, then $\sum_{C \in \mathcal{C}} |P_C| = O(n \log n)$.*

Proof. For every $0 \le i \le \log_2 n$, let $\mathcal{C}_i$ denote those curves $C \in \mathcal{C}$ for which $2^i \le |P_C| \le 2^{i+1}$. By Theorem 15, for every $0 \le i \le \log_2 n$, we have $\sum_{C \in \mathcal{C}_i} |P_C| = O((\frac{2^{i+1}}{2^i})^2 n) = O(n)$. Therefore, $\sum_{C \in \mathcal{C}} |P_C| = \sum_{i=0}^{\lfloor \log_2 n \rfloor} O(n) = O(n \log n)$. ∎

As an example, fix an integer $r \ge 2$. Let $G = (P, \mathcal{C})$ be a k-uniform topological hypergraph on n vertices. Assume that for any two curves $C_1, C_2 \in \mathcal{C}$, we have $|P_{C_1} \cap P_{C_2}| < r$. Then, clearly, for every r curves $C_1, \ldots, C_r \in \mathcal{C}$, we have $|P_{C_1} \cap \ldots \cap P_{C_r}| < r$. It follows now from Theorem 15 that

$$k|\mathcal{C}| = \sum_{C \in \mathcal{C}} |P_C| = O(\frac{k^2}{k^2} n) = O(n).$$

Hence, $|\mathcal{C}| = O(n/k)$. This roughly says that the sets in the family $\{P_C \mid C \in \mathcal{C}\}$, each having cardinality k, behave almost as if they were disjoint.

Acknowledgements We thank Eyal Ackerman, Kevin Buchin, and Christian Knauer for very helpful discussions We are grateful to Andrea Munaro for pointing out the *correct* formulation of Lemmma 3 . Research by Sarit Buzaglo and Rom Pinchasi was supported by the Israeli Science Foundation (grant no. 938/06).

References

1. P. Agarwal, E. Nevo, J. Pach, R. Pinchasi, M. Sharir, S. Smorodinsky, Lenses in arrangements of pseudocircles and their applications. J. ACM **51**, 139–186 (2004)
2. Ch. Chojnacki (A. Hanani), Über wesentlich unplättbare Kurven im dreidimensionalen Raume. Fund. Math. **23**, 135–142 (1934)
3. E. Helly, Über Systeme abgeschlossener Mengen mit gemeinschaftlichen Punkten. Monatshefte Mathematik **37**, 281–302 (1930)
4. P. Hall, On representatives of subsets. J. Lond. Math. Soc. **10**, 26–30 (1935)
5. J. Molnár, Über eine verallgemeinerung auf die kugelfläche eines topologischen satzes von Helly. Acta Math. Acad. Sci. **7**, 107–108 (1956)
6. J. Pach, G. Tóth, Disjoint edges in topological graphs, in *Combinatorial Geometry and Graph Theory*. Lecture Notes in Computer Science, vol. 3330 (Springer, Berlin, 2005), pp. 133–140
7. N. Sauer, On the density of families of sets. J. Comb. Theor. Ser. A **25**, 80–83 (1972)
8. S. Shelah, A combinatorial problem, stability and order for models and theories in infinite languages. Pacific J. Math. **41**, 247–261 (1972)
9. W.T. Tutte, Toward a theory of crossing numbers. J. Comb. Theor. **8**, 45–53 (1970)
10. V.N. Vapnik, A.a. Chervonenkis, On the uniform convergence of relative frequences of events to their probabilities. Theor. Prob. Appl. **16**, 264–280 (1971)

On Edge-Disjoint Empty Triangles of Point Sets

Javier Cano, Luis F. Barba, Toshinori Sakai, and Jorge Urrutia

Abstract Let P be a set of points in the plane in general position. Any three points $x,y,z \in P$ determine a triangle $\Delta(x,y,z)$ of the plane. We say that $\Delta(x,y,z)$ is empty if its interior contains no element of P. In this chapter, we study the following problems: What is the size of the largest family of edge-disjoint triangles of a point set? How many triangulations of P are needed to cover all the empty triangles of P? We also study the following problem: What is the largest number of edge-disjoint triangles of P containing a point q of the plane in their interior? We establish upper and lower bounds for these problems.

1 Introduction

Let P be a set of n points in the plane in general position. A geometric graph on P is a graph G whose vertices are the elements of P, two of which are adjacent if they are joined by a straight-line segment. We say that G is a plane if it has no edges that cross each other. A triangle of G consists of three points $x,y,z \in P$ such that xy, yz, and zx are edges of G; we will denote it as $\Delta(x,y,z)$. If, in addition, $\Delta(x,y,z)$ contains no elements of P in its interior, we say that it is empty.

J. Cano • L.F. Barba (✉)
Posgrado en Ciencia e Ingeniería de la Computación, Universidad Nacional Autónoma de México, México, D.F., México
e-mail: j_cano@uxmcc2.iimas.unam.mx; l.barba@uxmcc2.iimas.unam.mx

J. Urrutia
Instituto de Matemáticas, Universidad Nacional Autónoma de México, México, D.F., México.
Partially supported by CONACYT of México grant number: CB-2012-01-0178379
e-mail: urrutia@matem.unam.mx

T. Sakai
Research Institute of Educational Development, Tokai University, Kanagawa, Japan
e-mail: sakai@tokai-u.jp

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_7,

In a similar way, we say that if $x, y, z \in P$, then $\Delta(x, y, z)$ is a *triangle* of P, and that xy, yz, and zx are the edges of $\Delta(x, y, z)$. If $\Delta(x, y, z)$ is empty, it is called a 3-*hole* of P. A 3-hole of P can be thought of as an empty triangle of the complete geometric graph $\mathcal{K}_P$ on P. We remark that $\Delta(x, y, z)$ will denote a triangle of a geometric graph and also a triangle of a point set.

A well-known result in graph theory says that for $n = 6k+1$ or $n = 6k+3$, the edges of the complete graph K_n on n vertices can be decomposed into a set of $\binom{n}{2}/3$ edge-disjoint triangles. These decompositions are known as Steiner triple systems [23]; see also Kirkman's schoolgirl problem [17, 22]. In this chapter, we address some variants of that problem, but for geometric graphs.

Given a point set P, let $\delta(P)$ be the size of the largest set of edge-disjoint empty triangles of P. It is easy to see that for point sets in convex position with $n = 6k+1$ or $n = 6k+3$ elements, $\delta(P) = \binom{n}{2}/3$. Indeed, any triangle of P is empty, and the problem is the same as that of decomposing the edges of the complete geometric graph $\mathcal{K}(P)$ on P into edge-disjoint triangles. On the other hand, we prove that for some point sets, namely Horton point sets, $\delta(P)$ is $O(n \log n)$.

We then study the problem of covering the empty triangles of point sets with as few triangulations of P as possible. For point sets in convex position, we prove that we need essentially $\binom{n}{3}/4$ triangulations; our bound is tight. We also show that there are point sets P for which $O(n \log n)$ triangulations are sufficient to cover all the empty triangles of P for a given point set P.

Finally, we consider the problem of finding a point q not in P contained in the interior of many edge-disjoint triangles of P. We prove that for any point set, there is a point $q \notin P$ contained in at least $n^2/12$ edge-disjoint triangles. Furthermore, any point in the plane, not in P, is contained in at most $n^2/9$ edge-disjoint triangles of P, and this bound is sharp. In particular, we show that this bound is attained when P is the set of vertices of a regular polygon.

1.1 Preliminary Work

The study of counting and finding k-holes in point sets has been an active area of research since Erdős and Szekeres [11, 12] asked about the existence of k-holes in planar point sets. It is known that any point set with at least 10 points contains 5-holes; e.g., see [14]. Horton [15] proved that for $k \geq 7$, there are point sets containing no k-holes. The question of the existence of 6-holes remained open for many years, but recently Nicolás [19] proved that any point set with sufficiently many points contains a 6-hole. A second proof of this result was subsequently given by Gerken [13].

The study of properties of the set of triangles generated by point sets on the plane has been of interest for many years. Let $f_k(n)$ be the minimum number of k-holes that a point set has. Katchalski and Meir [16] proved that $\binom{n}{2} \leq f_3(n) \leq cn^2$ for some $c < 200$; see also Purdy [21]. Their lower bounds were improved by Dehnhardt [9] to $n^2 - 5n + 10 \leq f_3(n)$. He also proved that $\binom{n-3}{2} + 6 \leq f_4(n)$. Point sets with few k-holes for $3 \leq k \leq 6$ were obtained by Bárány and Valtr [2]. The interested reader can read [18] for a more accurate picture of the developments in this area of research.

Chromatic variants of the Erdős–Szekeres problem have recently been studied by Devillers, Hurtado, Károly, and Seara [10]. They proved among other results that any bichromatic point set contains at least $\frac{n}{4} - 2$ compatible monochromatic empty triangles. Aichholzer et al. [1] proved that any bichromatic point set always contains $\Omega(n^{5/4})$ empty monochromatic triangles; this bound was improved by Pach and Tóth [20] to $\Omega(n^{4/3})$.

2 Sets of Edge-Disjoint Empty Triangles in Point Sets

Let P be a set of points in the plane, and let $\delta(P)$ be the size of the largest set of edge-disjoint empty triangles of the complete graph $\mathcal{K}(P)$ on P. In this section we study the following problem:

Problem 1. How small can $\delta(P)$ be?

We show that if P is a Horton set, then $\delta(P)$ is $O(n \log n)$. By Kirkman's result, for points in convex position with $n = 6k+1$ and $n = 6k+3$, $\delta(P)$ is $\frac{\binom{n}{3}}{3}$.

For any integer $k \geq 1$, Horton [15] recursively constructed a family of point sets H_k of size 2^k as follows:

(a) $H_1 = \{(0,0),(1,0)\}$.
(b) H_k consists of two subsets of points H_{k-1}^- and H_{k-1}^+ obtained from H_{k-1} as follows: If $p = (i,j) \in H_{k-1}$, then $p' = (2i, j) \in H_{k-1}^-$ and $p'' = (2i+1, j+d_k) \in$

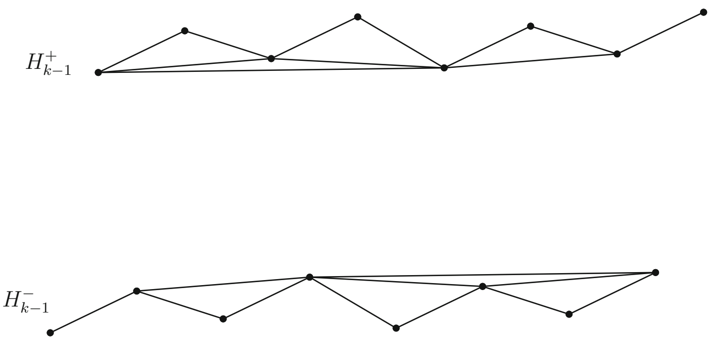


Fig. 1 H_4. The edges of H_3^+ (respectively, H_3^-) visible from *below* (respectively, *above*), are shown

H^+_{k-1}. The value d_k is chosen large enough such that any line ℓ passing through two points of H^+_{k-1} leaves all the points of H^-_{k-1} below it; see Fig. 1.

We say that a line segment pq joining two elements p and q of H_k is *visible from below* (respectively, *above*) if there is no point of H_k below it (respectively, above it); that is there is no element r of H_k such that the vertical line through r intersects pq above r (respectively, below r). Let $B(H_k)$ be the set of line segments of H_k visible from below. The following result, which we will use later, was proved by Bárány and Valtr in [2]; see also [3]:

Lemma 1. $|B(H_k)| = 2^{k+1} - (k+2)$.

The following result is proved in [3] by using this lemma:

Theorem 1. *For every $n = 2^k$, $k \geq 1$, there is a point set (namely, H_k) such that there is a geometric graph on H_k with $\binom{n}{2} - O(n \log n)$ edges with no empty triangles.*

In other words, it is always possible to remove $O(n \log n)$ edges from the complete graph $\mathcal{K}_{H_k}$ in such a way that the remaining graph contains no empty triangles. The main idea is that by removing from $\mathcal{K}_{H_k}$ all the edges of H^+_{k-1} (respectively, H^-_{k-1}) visible from below (respectively, above), no empty triangle remains with vertices in both H^+_{k-1} and H^-_{k-1}.

Observe now that if a geometric graph has k edge-disjoint empty triangles, then we need to take at least k edges away from G for the graph that remains to contain no empty triangles. It follows now that the complete graph $\mathcal{K}_{H_k}$ has at most $O(n \log n)$ edge-disjoint empty triangles. Thus, we have proved

Theorem 2. *There is a point set, namely, H_k, such that any set of edge-disjoint empty triangles of H_k contains at most $O(n \log n)$ elements.*

Clearly, for any point set P, the size of the largest set of edge-disjoint triangles of P is at least linear. We conjecture

Conjecture 1. Any point set P in general position always contains a set with at least $O(n \log n)$ edge-disjoint empty triangles.

3 Covering the Triangles of Point Sets with Triangulations

An empty triangle t of a point set P is covered by a triangulation T of P if one of the faces of T is t. In this section, we consider the following problem:

Problem 2. How many triangulations of a point set are needed such that each empty triangle of P is covered by at least one triangulation?

This problem, which is interesting in its own right, will help us in finding point sets for which $\delta(P)$ is large. We start by studying Problem 2 for point sets in convex position, and then for point sets in general position.

3.1 Points in Convex Position

All point sets P considered in this subsection will be assumed to be in convex position, and their elements labeled $\{p_0,\ldots,p_{n-1}\}$ in counterclockwise order around the boundary of $\mathrm{CH}(P)$. Since any triangulation of a point set of n points in convex position corresponds to a triangulation of a regular polygon with n vertices, solving Problem 2 for point sets in convex position is equivalent to solving it for point sets whose elements are the vertices of a regular polygon. Suppose then that P is the set of vertices of a regular polygon and that c is the center of such a polygon.

A triangle is called an *acute* triangle if all of its angles are smaller than $\frac{\pi}{2}$. We recall the following result in elementary geometry given without proof.

Observation 1. *A triangle with vertices in P is acute if and only if it contains c in its interior.*

The following result is relatively well known.

Lemma 2. *Let P be the set of vertices of a regular n-gon Q and c the center of Q. Then*

- *If n is even, c is contained in the interior of $\frac{1}{4}\left[\binom{n}{3}-\frac{n(n-2)}{2}\right]$ acute triangles of P.*
- *If n is odd, c is contained in $\left[\binom{n}{3}-\frac{n(n-1)(n-3)}{8}\right]=\frac{1}{4}\left[\binom{n}{3}+\frac{n(n-1)}{2}\right]$ acute triangles of P.*

Let $f(n)=\frac{1}{4}\left[\binom{n}{3}+\frac{n(n-2)}{2}\right]$ for n even and $f(n)=\frac{1}{4}\left[\binom{n}{3}+\frac{n(n-1)}{2}\right]$ for n odd. We now prove

Theorem 3. *$f(n)$ triangulations are always sufficient, and always necessary, to cover all the triangles of a regular polygon.*

Proof. Suppose first that n is even. For each vertex p_i of P, let $\alpha(p_i)=p_{i+\frac{n}{2}}$ be the antipodal vertex of p_i in P, where addition is taken mod n. Suppose that $\Delta(p_i,p_j,p_k)$ is an acute triangle of P (i.e., it contains c in its interior), $i<j<k$. Let $t_4(i,j,k)$ be the following set of four triangles:

$$t_4(i,j,k)=\{\Delta(p_i,p_j,p_k),\Delta(\alpha(p_i),p_j,p_k),\Delta(p_i,\alpha(p_j),p_k),\Delta(p_i,p_j,\alpha(p_k))\};$$

see Fig. 2a.

It is easy to see that all the triangles of P except those that have a right angle are in

$$\bigcup t_4(i,j,k),$$

where i,j,k range over all triples such that $\Delta(p_i,p_j,p_k)$ contains c in its interior.

On the other hand, it is easy to see that if a triangle t of P contains c in the middle of one of its edges (clearly, t is a right triangle), this edge joins two antipodal vertices of P; see Fig. 2b). Thus, we have exactly

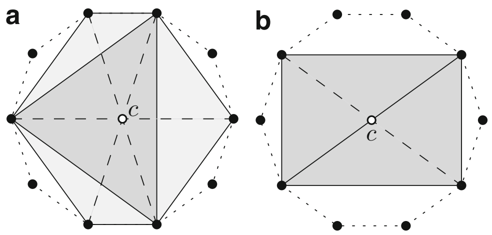


Fig. 2 (**a**) Constructing $t_4(i,j,k)$, and (**b**) pairing triangles sharing an edge, which contains c in the middle

$$\frac{n}{2} \times (n-2)$$

such triangles. It is easy to find

$$\frac{n(n-2)}{4}$$

triangulations of P such that each of them cover two of these triangles. Since each triangulation of P contains exactly one acute triangle of P or two triangles sharing an edge that contains c at its middle point, it follows that

$$\frac{1}{4}\left[\binom{n}{3} - \frac{n(n-2)}{2}\right] + \frac{n(n-2)}{4} = \frac{1}{4}\left[\binom{n}{3} + \frac{n(n-2)}{2}\right]$$

triangulations are necessary and sufficient to cover all the triangles of P. To show that this number of triangulations is needed, we point out that any two acute triangles of P cannot belong to the same triangulation (note that they intersect at c). Moreover, these triangulations are different from those containing right triangles. Our result follows.

A similar argument follows for n odd, except that some extra care has to be paid to the way in which we group the nonacute triangles of P around the acute triangles of P. □

Thus, the number of triangulations needed to cover all the triangles of P is asymptotically $\binom{n}{3}/4$. The next result follows trivially.

Corollary 1. *Let P be a set of n points in convex position and p any point in the interior of $CH(P)$. Then p belongs to the interior of at most $\frac{\binom{n}{3}}{4} + O(n^2)$ triangles of P.*

3.2 Covering the Empty Triangles on the Horton Set

We will now show that all the empty triangles in H_k can be covered with $O(n \log n)$ triangulations. The bound is tight.

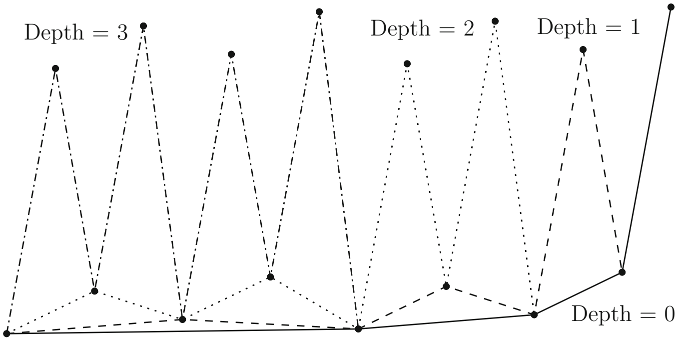


Fig. 3 The depth of an edge

Consider an edge e of H_k that is visible from below, and a vertical line ℓ that intersects e at a point q in the interior of e. The depth of e is the number of edges of H_k, visible from below, intersected by ℓ below q. It is not hard to see that the maximal depth of an edge of H_k visible from below is at most $\log n - 1$ and that this bound is tight; see Fig. 3. Moreover, it is easy to see that the union of all edges of H_k with the same depth is an x-monotone path. Now we can prove

Theorem 4. $\Theta(n \log n)$ *triangulations of* H_k *are necessary and sufficient to cover the set of empty triangles of* H_k.

Proof. Consider the sets H_{k-1}^{+} and H_{k-1}^{-}. We will show how to cover all the empty triangles of H_k with two vertices in H_{k-1}^{+} and one in H_{k-1}^{-} with $O(n \log n)$ triangulations. Label the elements of H_{k-1}^{-} from left to right as $p_0, \dots, p_{\frac{n}{2}-1}$.

For each $0 \le d \le k-1$, proceed as follows: For every $p_j \in H_{k-1}^{-}$, join p_j to the endpoints of all the edges of H_{k-1}^{+} of depth d. This gives us a set $ID_{d,j}^{+}$ of interior-disjoint empty triangles. It is not hard to see that if $(d, j) \neq (d', j')$, then $ID_{d,j}^{+} \cap ID_{d',j'}^{+} = \emptyset$.

It is easy to see that the union of these sets covers all the empty triangles with two vertices in H_{k-1}^{+} and one in H_{k-1}^{-}. In a similar way, cover all the triangles with two vertices in H_{k-1}^{-}, and one in H_{k-1}^{+} with a family of sets $ID_{d,j}^{-}$.

Let ℓ_1 be the straight line connecting the leftmost point in H_{k-1}^{+} to the rightmost point in H_{k-1}^{-}, and ℓ_2 the straight line that connects the rightmost point in H_{k-1}^{+} with the leftmost point of H_{k-1}^{-}. Let q be a point slightly above the intersection point of ℓ_1 with ℓ_2.

It is clear that for each $ID_{d,j}^{+}$ there is exactly one empty triangle that contains q in its interior. This implies that q is contained in $\Omega(n \log n)$ empty triangles, and thus $\Omega(n \log n)$ triangulations are necessary to cover all the empty triangles in H_k.

Now we show that $O(n \log n)$ of H_k triangulations are sufficient. Consider each set $ID_{d,j}^{+}$ and $ID_{d,j}^{-}$, and complete it to a triangulation. This gives us $O(n \log n)$ triangulations that cover all the triangles with vertices in both H_{k-1}^{+} and H_{k-1}^{-}.

Take a set of triangulations $\mathcal{T}_{k-1}^{+} = \{T_1^{+}, \ldots, T_m^{+}\}$ of H_{k-1}^{+} that covers all of its empty triangles. Since H_{k-1}^{+} and H_{k-1}^{-} are isomorphic, we can find a set of triangulations $\mathcal{T}_{k-1}^{-} = \{T_1^{-}, \ldots, T_m^{-}\}$ of H_{k-1}^{-} that covers all the empty triangles of H_{k-1}^{-} such that T_i^{+} is isomorphic to T_i^{-}. For each i, we can find a triangulation T_i of H_k that contains T_i^{+} and T_i^{-} as induced subgraphs.

Thus, if $T(n)$ is the number of triangulations required to cover the empty triangles of H_k, the following recurrence holds for $n = 2^k$:

$$T(n) = T\left(\frac{n}{2}\right) + O(n \log n).$$

This solves to $T(n) = O(n \log n)$, and our result follows. □

We conclude this section with the following conjecture.

Conjecture 2. At least $\Omega(n \log n)$ triangulations are needed to cover all the empty triangles of any point set with n points.

4 A Point in Many Edge-Disjoint Triangles

The problem of finding a point contained in many triangles of a point set was solved by Boros and Füredi [4]; see also Bukh [6]. They proved

Theorem 5. *For any set P of n points in general position, there is a point in the interior of the convex hull of P contained in $\frac{2}{9}\binom{n}{3} + O(n^2)$ triangles of P. The bound is tight.*

We now study a variant to this problem, in which we are interested in finding a point in many *edge-disjoint* triangles. We consider the following.

Problem 3. Let P be a set of points in the plane in general position, and $q \notin P$ a point of the plane. What is the largest number of edge-disjoint triangles of P such that q belongs to the interior of all of them?

We start by giving some preliminary results, and then we study Problem 3 for point sets in general position and sets of vertices of regular polygons.

Given a point set P, and a point q not in P, let $\mathcal{T}(P,q)$ [or $\mathcal{T}(q)$ for short] be the set of triangles of P that contain q. We define the graph $G(P,q)$ whose vertex set is $\mathcal{T}(q)$ in which two triangles are adjacent if they share an edge; see Fig. 4. We may assume that q does not belong to any line passing through two elements of P. We now prove

Lemma 3. *The degree of every vertex of $G(P,q)$ is exactly $n-3$.*

Proof. Let $\Delta(x,y,z)$ be a triangle that contains q in its interior. Let p be any point in $P \setminus \{x,y,z\}$. Then exactly one of the triangles $\Delta(x,y,p)$, $\Delta(x,p,z)$, or $\Delta(p,y,z)$

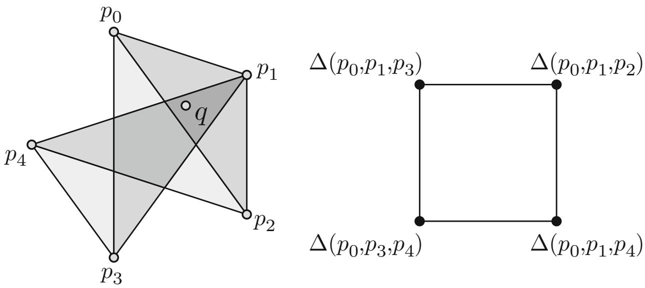
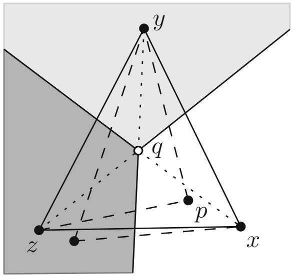


Fig. 4 $G(P,q)$

Fig. 5



contains q; see Fig. 5. That is, exactly one of $\Delta(x,y,p)$, $\Delta(x,p,z)$, or $\Delta(p,y,z)$ belongs to $\mathcal{T}(q)$. Our result follows. □

Observe now that finding sets of edge-disjoint triangles that contain q is equivalent to finding independent sets in $G(P,q)$. Let $\tau(P,q)$ (or $\tau(q)$ for short) be the largest number of edge-disjoint triangles on P containing q. We now prove

Lemma 4.

$$\frac{|\mathcal{T}(q)|}{n-2} \leq \tau(q) \leq \frac{3|\mathcal{T}(q)|}{n}.$$

Proof. It follows from Lemma 3 that the size of the largest independent set of $G(P,q)$ is at least $\frac{|\mathcal{T}(q)|}{n-2}$. To prove our upper bound, it is sufficient to observe that if a vertex of $G(P,q)$ is not in an independent set I of $G(P,q)$, then it is adjacent to at most three vertices in it, one per each of its edges. Hence, by counting the number of edges connecting a vertex in I to another in $\mathcal{T}(q) \setminus I$, we obtain

$$(n-3)|I| \leq 3|\mathcal{T}(q) \setminus I|.$$

Our result follows. □

From Theorem 5 and Lemma 4, it is easy to see that in any set of n points in general position on the plane, there is a point q such that

$$\frac{n^2}{27} + O(n) \approx \frac{\frac{2}{9}\binom{n}{3} + O(n^2)}{n-2} \leq \tau(q) \leq \frac{3 \cdot \frac{2}{9}\binom{n}{3} + O(n^2)}{n} \approx \frac{n^2}{9} + O(n).$$

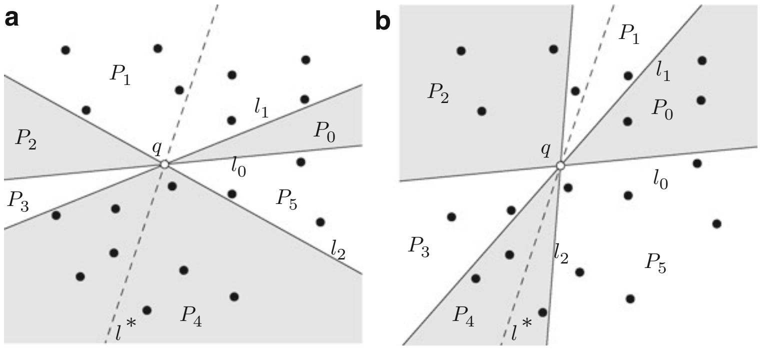


Fig. 6 Partitions of P

Thus, we have

Corollary 2. *For any point set in general position on the plane, there is a point q such that $\tau(q) \leq \frac{n^2}{9} + O(n)$.*

We now prove an even stronger result. We now prove

Proposition 1. *Let P a set of n points in general position on the plane. Then for any point $q \notin P$ of the plane, $\tau(q) \leq n^2/9$.*

Proof. Let $q \notin P$ be any point of the plane. If q is on a straight line passing through two elements of P, then by slightly moving it, q could be moved to a position in which it is contained in more edge-disjoint triangles. Thus, assume that q is not on any straight line through two elements of P.

First, we show the following lemma:

Lemma 5. *There exist three straight lines passing through q such that they partition P into six subsets $P_0, P_1, \ldots, P_5$ in counterclockwise order around q, with $|P_0| = |P_2| = |P_4|$ (we allow the possibility that $P_i = \emptyset$ for some i).*

Proof. Let l_0 be a straight line passing through q such that one of the half-planes bounded by l_0, say the one on top of it, contains an even number of elements of P. Take other straight lines l_1 and l_2 passing through q, and define the subsets P_i of P, $0 \leq i \leq 5$, as shown in Fig. 6a, where $|P_0 \cup P_1 \cup P_2|$ is even. Let l^* be a straight line passing through q, equipartitioning the elements of $P_0 \cup P_1 \cup P_2$.

Choose l_1 and l_2 such that initially $|P_0| = |P_2| = |P_3| = |P_5| = 0$. From their initial positions, rotate l_1 counterclockwise and l_2 clockwise around q in such a way that P_0 and P_2 always contain the same number of elements, and until they both reach the position of l^* at the same time, and the boundary of P_4 always contains no more than one element of P.

Initially, $|P_4| \geq 0 = |P_0|$. On the other hand, we have $|P_4| = 0 \leq |P_0|$ when l_1 and l_2 reach the position of l^*. Hence, at some point while rotating l_1 and l_2, we have that $|P_0| = |P_2| = |P_4|$; see Fig. 6b. □

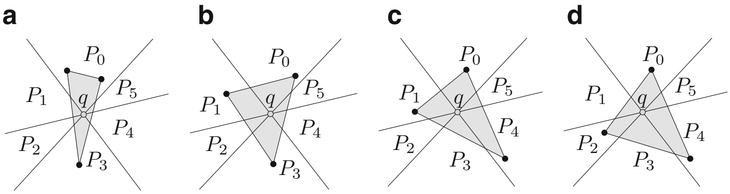


Fig. 7 Triangles in the $\mathcal{T}_{ijk}$'s

Let $P_0, P_1, \ldots, P_5$ be as in Lemma 5. Write $|P_i| = n_i$ for $0 \le i \le 5$ (we have $n_0 = n_2 = n_4$). We henceforth read indices modulo 6. Let $\mathcal{T}$ be a set of edge-disjoint triangles with vertices in P, containing q in its interior. For integers i, j, k, let $\mathcal{T}_{ijk}$ denote the set of elements of $\mathcal{T}$ such that it has one vertex in P_i, another in P_j and the other in P_k, and let $t_{ijk} = |\mathcal{T}_{ijk}|$; see Fig. 7.

Then

$$\mathcal{T} = \left[\cup_{i=0}^5 \mathcal{T}_{ii(i+3)}\right] \cup \left[\cup_{i=0}^5 \mathcal{T}_{i(i+1)(i+3)}\right] \cup \left[\cup_{i=0}^5 \mathcal{T}_{i(i+1)(i+4)}\right] \cup \left[\cup_{i=0}^5 \mathcal{T}_{i(i+2)(i+4)}\right]$$
$$= \left[\cup_{i=0}^5 \mathcal{T}_{ii(i+3)}\right] \cup \left[\cup_{i=0}^5 \mathcal{T}_{i(i+2)(i+5)}\right] \cup \left[\cup_{i=0}^5 \mathcal{T}_{i(i+2)(i+3)}\right] \cup \left[\cup_{i=0}^5 \mathcal{T}_{i(i+2)(i+4)}\right].$$

For integers i, j, let E_{ij} denote the set of all segments connecting an element of P_i and another of P_j. Then for each integer i, $|E_{i(i+2)}| = n_i n_{i+2}$ and $\mathcal{T}_{i(i+2)(i+3)} \cup \mathcal{T}_{i(i+2)(i+4)} \cup \mathcal{T}_{i(i+2)(i+5)}$ is the set of elements of $\mathcal{T}$ that has a side belonging to $E_{i(i+2)}$. Hence, we have

$$f(i) \equiv t_{i(i+2)(i+3)} + t_{i(i+2)(i+4)} + t_{i(i+2)(i+5)} \le n_i n_{i+2} \tag{1}$$

for each i. Similarly, by considering the cardinality of $E_{i(i+3)}$, we obtain

$$g(i) \equiv 2t_{ii(i+3)} + t_{i(i+1)(i+3)} + t_{i(i+2)(i+3)}$$
$$+2t_{i(i+3)(i+3)} + t_{i(i+3)(i+4)} + t_{i(i+3)(i+5)} \le n_i n_{i+3} \tag{2}$$

for each i. By (1) and (2), we have

$$\sum_{i=0}^{5} f(i) + 2\sum_{i=0}^{2} g(i) \le \sum_{i=0}^{5} n_i n_{i+2} + 2\sum_{i=0}^{2} n_i n_{i+3}. \tag{3}$$

Since $g(i) = (t_{i(i+2)(i+3)} + t_{j(j+2)(j+3)}) + (t_{j'(j'+2)(j'+5)} + t_{j''(j''+2)(j''+5)}) + 2(t_{ii(i+3)} + t_{jj(j+3)})$, where $j = i+3$, $j' = i+1$, $j'' = j'+3$,

$$\sum_{i=0}^{5} f(i) + 2\sum_{i=0}^{2} g(i) = \sum_{i=0}^{5} (t_{i(i+2)(i+3)} + t_{i(i+2)(i+4)} + t_{i(i+2)(i+5)})$$
$$+2\sum_{i=0}^{5} (t_{i(i+2)(i+3)} + t_{i(i+2)(i+5)}) + 4\sum_{i=0}^{5} t_{ii(i+3)}$$
$$= 3|\mathcal{T}| + \sum_{i=0}^{5} t_{ii(i+3)} \ \ge \ 3|\mathcal{T}|. \tag{4}$$

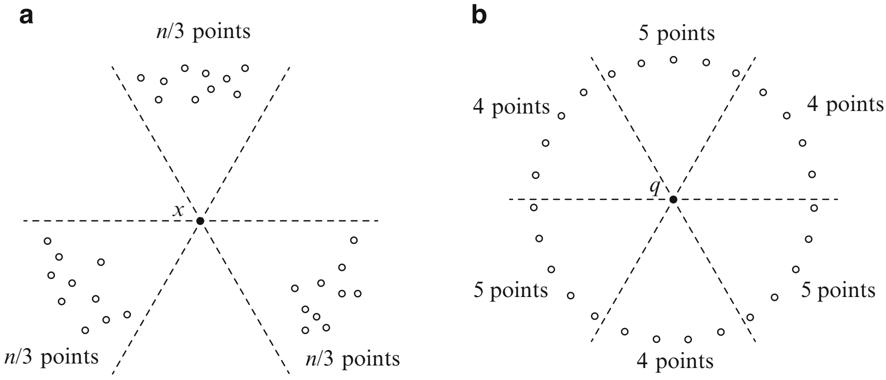


Fig. 8 A vertex set of a regular 27-gon

On the other hand, if we denote the right-hand side of (3) by S,

$$S = (n_0n_2 + n_2n_4 + n_4n_0) + (n_1n_3 + n_3n_5 + n_5n_1) + 2(n_0n_3 + n_2n_5 + n_4n_1) = \frac{l^2}{3} + \frac{2lm}{3} + (n_1n_3 + n_3n_5 + n_5n_1), \quad (5)$$

where $l = n_0 + n_2 + n_4$ (recall that $n_0 = n_2 = n_4$) and $m = n_1 + n_3 + n_5$. Since $n_1n_3 + n_3n_5 + n_5n_1 = [m^2 - (n_1^2 + n_3^2 + n_5^2)]/2$ and since $n_1^2 + n_3^2 + n_5^2 \geq m^2/3$ with equality if and only if $n_1 = n_3 = n_5$, we have $n_1n_3 + n_3n_5 + n_5n_1 \leq m^2/3$. From this and (5), it follows that

$$S \leq \frac{l^2}{3} + \frac{2lm}{3} + \frac{m^2}{3} = \frac{(l+m)^2}{3} = \frac{n^2}{3}. \quad (6)$$

Now combining (3), (4) and (6), we obtain $|\mathcal{T}| \leq n^2/9$, as desired. □

To achieve the equality, it is necessary that $n_0 = n_2 = n_4$ and $n_1 = n_3 = n_5$ for some partition (Fig. 8).

We now prove

Proposition 2. *Let n be a positive integer and P a set of n points in general position on the plane. Then there exists a point q on the plane such that* $\tau(q) \geq \frac{n^2}{12}$.

Proof. We use the following lemma, which was proved by Ceder [7] (see also [5]) and applied by Bukh [6] to obtain a lower bound of $\max_q |\mathcal{T}(q)|$ for given P:

Lemma 6. *There exist three straight lines such that they intersect at a point q and partition the plane into six open regions each of which contains at least* $n/6 - 1$ *elements of P.*

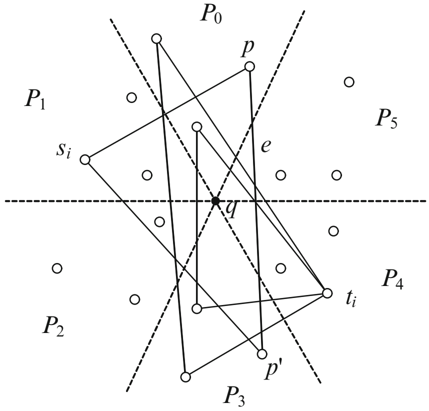


Fig. 9 Matching M_i (*bold lines*) and triangles using edges of M_i

Let q be as in Lemma 6. We may assume that q is not on any straight line passing through two elements of P. Let $m = \lceil n/6 \rceil - 1$ and denote by $D_0, D_1, \ldots, D_5$ the six regions in counterclockwise order around q. For each $0 \le i \le 5$, let P_i be a subset of $P \cap D_i$ with $|P_i| = m$; see Fig. 9.

Now consider the geometric complete bipartite graph with vertex set $P_0 \cup P_3$ and edge set $E = \{pp' \,|\, p \in P_0,\, p' \in P_3\}$. As a consequence of a well-known result in graph theory, E can be decomposed into m subsets M_i, $0 \le i \le m-1$, such that each M_i is a perfect matching, i.e., consisting of m independent edges. Let $P_1 = \{s_1, s_2, \ldots, s_m\}$ and $P_4 = \{t_1, t_2, \ldots, t_m\}$. For each i and each element $e = pp' \in M_i$, where $p \in P_0$ and $p' \in P_3$, let u_i denote either s_i or t_i according to whether $pp' \cap D_1 = \emptyset$ or $pp' \cap D_4 = \emptyset$. Then $\triangle(p, p', u_i)$ contains q in its interior. Observe that all of the m triangles in $\mathcal{T}_i = \{\triangle(p, p', u_i) \,|\, e = pp' \in M_i\}$ are edge-disjoint, and all of the m^2 triangles in $\mathcal{T}_{03} = \cup_{i=0}^{m} \mathcal{T}_i$ are edge-disjoint as well.

Define the sets $\mathcal{T}_{14}$ and $\mathcal{T}_{25}$ of triangles similarly (the elements of $\mathcal{T}_{14}$ are triangles with one vertex in P_1, another in P_4, and the other in $P_2 \cup P_5$, while the elements of $\mathcal{T}_{25}$ are triangles with one vertex in P_2, another in P_5, and the other in $P_3 \cup P_0$). It can be observed that all of the $3m^2 = n^2/12 - O(n)$ triangles in $\mathcal{T}_{03} \cup \mathcal{T}_{14} \cup \mathcal{T}_{25}$ are edge-disjoint. □

Thus by using Corollary 2, Proposition 1, and Proposition 2, we have

Theorem 6. *In any point set in general position, there is a point q such that $\frac{n^2}{12} \le \tau(q) \le \frac{n^2}{9}$. Moreover, for any q, $\tau(q) \le \frac{n^2}{9}$.*

4.1 Regular Polygons

By Theorem 6, any point in the interior of the convex hull of a point set is contained in at most $n^2/9$ edge-disjoint triangles of P. It is also easy to construct point sets for which that bound is tight; see Fig. 8a). In fact, the point sets in that figure can be chosen in convex position.

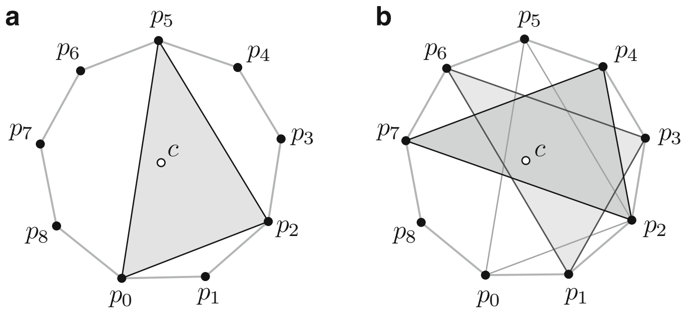


Fig. 10 (**a**) The triple $(1,2,3)$, and p_0 determine $\Delta(p_0, p_2, p_5)$. (**b**) $S(1,2,3)$ is obtained by rotating $\Delta(p_0, p_2, p_5)$, obtaining a set of 9 edge-disjoint triangles

We now show that the bound in Theorem 6 is also achieved when P is the set of vertices of a regular polygon. We found proving this result to be a challenging problem. In what follows, we will assume that $n = 9m$, $m \geq 1$.

Let (a_i, b_i, c_i) be an ordered set of integers. We call (a_i, b_i, c_i) a *triangular triple* if it satisfies the following conditions:

(a) a_i, b_i, and c_i are all different,
(b) $a_i + b_i + c_i = n - 3$, and
(c) $1 \leq a_i, b_i, c_i \leq \frac{n-3}{2}$.

Observe that for any vertex p_r of P, a triangular triple (a_i, b_i, c_i), defines a triangle $\Delta(p_r, p_{r+a_i+1}, p_{r+a_i+b_i+2})$ of P. Moreover, condition c) above ensures that $\Delta(p_r, p_{r+a_i+1}, p_{r+a_i+b_i+2})$ is acute, and hence it contains the center c of P. Note that since $a_i + b_i + c_i = n - 3$, $p_r = p_{r+a_i+b_i+c_i+3}$, addition taken mod n. Thus, the edges of $\Delta(p_r, p_{r+a_i+1}, p_{r+a_i+b_i+2})$ skip a_i, b_i, and c_i vertices of P, respectively; see Fig. 10a.

Let $S(a_i, b_i, c_i) = \{\Delta(p_r, p_{r+a_i+1}, p_{r+a_i+b_i+2}) : p_r \in P\}$. The set $S(a_i, b_i, c_i)$ can be seen as the set of triangles obtained by rotating $\Delta(p_0, p_{0+a_i+1}, p_{0+a_i+b_i+2})$ around the center of P; see Fig. 10b. The next observation will be useful.

Observation 2. *Let (a_i, b_i, c_i) and (a_j, b_j, c_j) be triangular triples of P such that $\{a_i, b_i, c_i\} \cap \{a_j, b_j, c_j\} = \emptyset$, $i \neq j$. Then all of the triangles in $S(a_i, b_i, c_i) \cup S(a_j, b_j, c_j)$ are edge-disjoint.*

Consider a set $C = \{(a_0, b_0, c_0), \ldots, (a_{k-1}, b_{k-1}, c_{k-1})\}$ of ordered triangular triples. We say that C is a *generating set* of triangular triples if the following condition holds:

$$\{a_i, b_i, c_i\} \cap \{a_j, b_j, c_j\} = \emptyset,\ i \neq j.$$

$(4,8,12)$	$(7,15,20)$	$(10,22,28)$	$(13,29,36)$	$(16,36,44)$
$(5,9,10)$	$(8,13,21)$	$(11,20,29)$	$(14,27,37)$	$(17,34,45)$
$(6,7,11)$	$(9,16,17)$	$(12,18,30)$	$(15,25,38)$	$(18,32,46)$
	$(10,14,18)$	$(13,23,24)$	$(16,23,39)$	$(19,30,47)$
	$(11,12,19)$	$(14,21,25)$	$(17,30,31)$	$(20,28,48)$
		$(15,19,26)$	$(18,28,32)$	$(21,37,38)$
		$(16,17,27)$	$(19,26,33)$	$(22,35,39)$
			$(20,24,34)$	$(23,33,40)$
			$(21,22,35)$	$(24,31,41)$
				$(25,29,42)$
				$(26,27,43)$

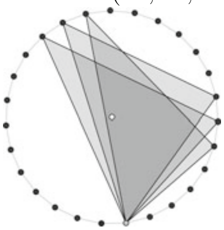


Fig. 11 Triangular triples for $n = 27, 45, 63, 81$ and 99

Note that $|S(a_i,b_i,c_i)| = n$, and thus

$$\bigcup_{(a_i,b_i,c_i)\in C} S(a_i,b_i,c_i)$$

contains nk edge-disjoint triangles containing the center P. Our task is now that of finding a generating set of as many triangular triples as possible.

Theorem 7. *Let P be the set of vertices of a regular polygon with $n = 9m$ vertices, and let c be its center. Then if m is odd, then $|\tau(c)| \geq \frac{n^2}{9}$, and if m is even, then $|\tau(c)| \geq \frac{n^2}{9} - n$.*

Proof. The proof when m is odd proceeds by constructing a generating set C with $\frac{n}{9}$ triangular triples. Let $k = \frac{9m-3}{6}$ and $k' = k + 2m - 1$. Given $i \in \{0,1,\ldots,m-1\}$, we define the ith ordered triple (a_i,b_i,c_i) as follows (see Fig. 11):

$$a_i = k + i,$$
$$b_i = \begin{cases} k' - 2i - 1 & \text{if} \quad i < \frac{m-1}{2}, \\ k' - 2i + m - 1 & \text{if} \quad i \geq \frac{m-1}{2}, \end{cases}$$
$$c_i = \begin{cases} k' + i + 1 + \frac{m+1}{2} & \text{if} \quad i < \frac{m-1}{2}, \\ k' + i + 1 - \frac{m-1}{2} & \text{if} \quad i \geq \frac{m-1}{2}. \end{cases}$$

We now prove that the triples (a_i,b_i,c_i) are triangular; that is, $a_i + b_i + c_i = n - 3$. Since $b_i + c_i = 2k' - i + \frac{m+1}{2}$ for all i,

$$a_i + b_i + c_i = k + 2k' + \frac{m+1}{2} = 9m - 3.$$

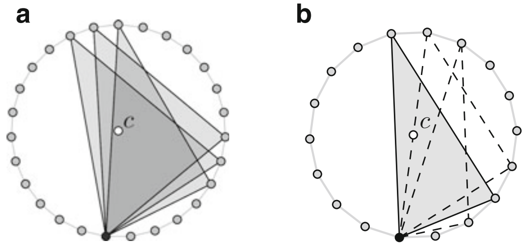


Fig. 12 (**a**) Triangular triples (a_i, b_i, c_i) for $n = 9 \cdot 3 = 27$ and (**b**) triples $(a'_i, b'_i, c'_i) = (a_i - 3, b_i - 3, c_i - 3)$ for $n = 9 \cdot 2 = 18$

It is easy to see that

$$\begin{aligned} k \leq\ & a_i \leq k + m - 1, \\ k + m = k' - m + 1 \leq\ & b_i \leq k', \\ k' + 1 \leq\ & c_i. \end{aligned}$$

Therefore, $a_i < b_j < c_k$ for every i, j, k. Also, by construction it can be verified that $a_i \neq a_j$, $b_i \neq b_j$, and $c_i \neq c_j$ for every $i \neq j$.

$$\text{Thus, the set } \bigcup_{(a_i, b_i, c_i) \in C} \{a_i, b_i, c_i\} \text{ contains no repeated elements.}$$

Finally, note that the maximum value that can be reached by c_i occurs when $i = \frac{m-3}{2}$, and thus,

$$c_i \leq k' + 1 + \frac{m-3}{2} + \frac{m+1}{2} = k' + m = \frac{9m-3}{2}.$$

Therefore, C is a generating set of triangular triples. Thus, c is contained in at least $\frac{n^2}{9}$ edge-disjoint triangles.

The proof when m is even proceeds by also constructing a set of m triples. We use the set of triples just constructed for $m+1$ and modify it as follows: Suppose that the set of $m+1$ triples is $\{(a_0, b_0, c_0), \ldots, (a_m, b_m, c_m)\}$.

Let $a'_i = a_i - 3$, $b'_i = b_i - 3$, and $c'_i = c_i - 3$, and consider $C' = \{(a'_i, b'_i, c'_i) \mid 0 \leq i \leq m\}$. C' induces a set of triangles in P. Nevertheless, $2n$ triangles do not contain the point c in their interior; see Fig. 12. Therefore, this construction guarantees that c is contained in at least $(m+1)n - 2n = \frac{n^2}{9} - n$ edge-disjoint triangles. □

5 A Point in Many Edge-Disjoint Empty Triangles

We conclude our chapter by briefly studying the problem of the existence of a point contained in many edge-disjoint empty triangles of a point set. We point out that if we are interested only in empty triangles containing a point, it is easy to see that for any point set P, there is always a point q contained in a linear number of (not necessarily edge-disjoint) empty triangles. This follows directly from the following facts:

1. Any point set P with n elements always determines at least a quadratic number of empty triangles [2, 16].
2. We can always choose $2n-c-2$ points in the plane such that any empty triangle of P contains one of them, where c is the number of vertices of the convex hull of P; see [8, 16].

We now prove

Theorem 8. *There are point sets P such that every $q \notin P$ is contained in at most a linear number of empty edge-disjoint triangles of P.*

Proof. Let H_k, H_{k-1}^+, and H_{k-1}^- be as defined in Sect. 2. Consider any set T_k^+ (respectively, T_k^-) of empty edge-disjoint triangles such that each of them has two vertices in H_{k-1}^+ (respectively, H_{k-1}^-) and the other in H_{k-1}^- (respectively, H_{k-1}^+). Let $t \in T_k^+$. Then the edge of t with both endpoints in H_{k-1}^+ is an edge of H_{k-1}^+ visible from below. Since the triangles in T_k^+ are edge-disjoint, the number of elements of T_k^+ is at most the number of edges of H_{k-1}^+ visible from below, which is a linear function in n. Thus, $|T_k^+| \in O(n)$. Similarly, we can prove that $|T_k^-| \in O(n)$.

Consider a point $q \in \mathrm{CH}(H_k) \setminus \mathrm{CH}(H_{k-1}^+) \cup \mathrm{CH}(H_{k-1}^-)$. Clearly, any empty triangle containing q belongs to some $T_k^+ \cup T_k^-$, and thus it belongs to at most a linear number of edge-disjoint triangles of H_k.

Suppose next that $q \in \mathrm{CH}(H_{k-1}^+) \cup \mathrm{CH}(H_{k-1}^-)$. Suppose without loss of generality that $q \in \mathrm{CH}(H_{k-1}^+)$ and that q belongs to a set S of edge-disjoint triangles of H_k. S may contain some triangles with vertices in both of H_{k-1}^+ and H_{k-1}^-. There are at most a linear number of such triangles. The remaining elements in S have all of their vertices in H_{k-1}^+. Thus, the number of edge-disjoint triangles containing q satisfies

$$T(n) \le T\left(\frac{n}{2}\right) + \Theta(n),$$

and thus q belongs to at most a linear number of edge-disjoint triangles.

The first part of our result follows. To show that our bound is tight, let q be as in the proof of Theorem 4. It is easy to see that q belongs to a linear number of triangles with vertices in both H_k^+ and H_k^-, and our result follows. □

We conclude with the following.

Conjecture 3. Let P be a set of n points in general position on the plane. Then there is always a point $q \notin P$ on the plane such that it is contained in at least $\log n$ edge-disjoint triangles of P.

Acknowledgements Our work was partially supported by projects MTM2009-07242, MTM2006-03909 (Spain), and SEP-CONACYT of México, Proyecto 80268.

References

1. O. Aichholzer, R. Fabila-Monroy, D. Flores-Pe-aloza, T. Hackl, C. Huemer, J. Urrutia, Empty monochromatic triangles. Comput. Geom. Theor. Appl. **42**, 934–938 (2009)
2. I. Bárány, P. Valtr, Planar point sets with a small number of empty convex polygons. Stud. Sci. Math. Hung. **41**(2), 243–266 (2004)
3. C. Bautista-Santiago, M. Heredia, C. Huemer, A. Ramírez-Vigueras, C. Seara, J. Urrutia, On the number of edges in geometric graphs without empty triangles (January 2011) to appear in Graphs and Combinatorics. DOI : 10.1007/s00373-012-1220-9
4. E. Boros, Z. Füredi, The number of triangles covering the center of an *n*-set. Geom. Dedicata **17**, 69–77 (1984)
5. R. Buck, E. Buck, Equipartitions of convex sets. Math. Mag. **22**, 195–198 (1949)
6. B. Bukh, A point in many triangles. Electron. J. Comb. **13**(10) (2006)
7. J. Ceder, Generalized sixpartite problems. Bol. Soc. Math. Mexicana **2**, 28–32 (1964)
8. J. Czyzowicz, E. Kranakis, J. Urrutia, Guarding the convex subsets of a point set, in *Proceedings of the 12th Canadian Conference on Computational Geometry*, 2000, New Brunswick, Canada, pp. 47–50
9. K. Dehnhardt, Leere konvexe Vielecke in ebenen Punktmengen. Dissertation, TU Braunschweig, 1987
10. O. Devillers, F. Hurtado, G. Károlyi, C. Seara, Chromatic variants of the Erdős–Szekeres theorem. Comput. Geom. Theor. Appl. **26**(3), 193–208 (2003)
11. P. Erdős, Some more problems on elementary geometry. Aust. Math. Soc. Gaz. **5**, 52–54 (1978)
12. P. Erdős, G. Szekeres, A combinatorial problem in geometry. Compos. Math. **2**, 463–470 (1935)
13. T. Gerken, Empty convex hexagons in planar point sets. Discr. Comput. Geom. **39**(1–3), 239–272 (2008)
14. H. Harborth, Konvexe Fünfecke in ebenen Punktmengen. Elem. Math. **33**, 116–118 (1978)
15. J.D. Horton, Sets with no empty convex 7-gons. Can. Math. Bull. **26**, 482–484 (1983)
16. M. Katchalski, A. Meir, On empty triangles determined by points in the plane. Acta Math. Hung. **51**, 323–328 (1988)
17. T. Kirkman, On a problem in combinatorics. Camb. Dublin Math. J. **2**, 191–204 (1847)
18. W. Morris, V. Soltan, The Erdős–Szekeres problem on points in convex position—a survey. Bull. (New Ser.) Am. Math. Soc. **37**(4), 437–458 (2000)
19. C.M. Nicolás, The empty hexagon theorem. Discr. Comput. Geom. **38**, 389–397 (2007)
20. J. Pach, G. Tóth, Monochromatic empty triangles in two-colored point sets, in *Geometry, Games, Graphs and Education: the Joe Malkevitch Festschrift* (COMAP, Bedford, 2008), pp. 195–198
21. G. Purdy, The minimum number of empty triangles. AMS Abstr. **3**, 318 (1982)
22. D. Ray-Chaudhuri, R. Wilson, Solution to Kirkman's schoolgirl problem. Proc. Symp. Pure Math. (Am. Math. Soc.) **19**, 187–204 (1971)
23. H.J. Ryser, *Combinatorial Mathematics* (Math. Assoc. Amer., Buffalo, 1963), pp. 99–102

Universal Sets for Straight-Line Embeddings of Bicolored Graphs

Josef Cibulka, Jan Kynčl, Viola Mészáros, Rudolf Stolař, and Pavel Valtr

Abstract A set S of n points is 2-*color universal* for a graph G on n vertices if, for every proper 2-coloring of G and for every 2-coloring of S with the same sizes of color classes as G, G is straight-line embeddable on S.

We show that the so-called double-chain is 2-color universal for paths if each of the two chains contains at least one fifth of all the points, but not if one of the chains is more than approximately 28 times longer than the other.

A 2-coloring of G is *equitable* if the sizes of the color classes differ by at most 1. A bipartite graph is *equitable* if it admits an equitable proper coloring. We study the case when S is the double-chain with chain sizes differing by at most 1 and G is an equitable bipartite graph. We prove that this S is not 2-color universal if G is not a forest of caterpillars and that it is 2-color universal for equitable caterpillars with at most one half nonleaf vertices. We also show that if this S is equitably 2-colored, then equitably properly 2-colored forests of stars can be embedded on it.

J. Cibulka (✉) • Rudolf Stolař
Department of Applied Mathematics, Charles University, Malostranské nám. 25, 11800 Prague, Czech Republic
e-mail: cibulka@kam.mff.cuni.cz; ruda@kam.mff.cuni.cz

J. Kynčl • V. Mészáros • P. Valtr
Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Malostranské nám. 25, 11800 Prague, Czech Republic
e-mail: kyncl@kam.mff.cuni.cz

V. Mészáros
Institute of Mathematics, Technical University of Berlin, Strasses des 17. Juni 136, 10623 Berlin, Germany
e-mail: meszaros@math.tu-berlin.de

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_8,

1 Introduction

1.1 Previous Results

It is frequently asked in geometric graph theory whether a given graph G can be drawn without edge crossings on a given planar point set S under some additional constraints on the drawing. In this chapter, we always assume $|V(G)| = |S|$.

One possibility is to prescribe a fixed position for each vertex of G. If the edges are allowed to be arbitrary curves, then we can obtain a planar drawing of an arbitrary graph G by moving vertices from an arbitrary planar drawing of G to the given points. Pach and Wenger showed [17] that every planar G with prescribed vertex positions can be drawn so that each edge is a piecewise-linear curve with $O(n)$ bends and that this bound is tight even if G is a path.

In another setting, we are only given the graph G and the set of points S, but we are allowed to choose the mapping between $V(G)$ and S. Kaufmann and Wiese [15] showed that two bends per edge are then always enough and for some graphs necessary. An *outerplanar graph* is a graph admitting a planar drawing where one face contains all vertices. By a result of Gritzmann et al. [11], outerplanar graphs are exactly the graphs with a straight-line planar drawing on an arbitrary set of points in general position.

In this chapter, we are dealing with a combination of these two versions. The vertices and points are colored with two colors and each vertex has to be placed on a point of the same color. An obvious necessary condition to find such a drawing is that each color class has the same number of vertices as points. We then say that the 2-coloring of S is *compatible* with the 2-coloring of $V(G)$.

A *caterpillar* is a tree in which the nonleaf vertices induce a path. The coloring of $V(G)$ is *proper* if it doesn't create any monochromatic edge. It is known that drawing some bicolored planar graphs on some bicolored point sets requires at least $\Omega(n)$ bends per edge [9], but caterpillars can be drawn with two bends per edge [9], and one bend per edge is enough for paths [10] and properly colored caterpillars [9].

We restrict our attention to the proper 2-colorings of a bipartite graph G. Then the question of the embeddability of G on a bicolored point set is very similar to finding a noncrossing copy of G in a complete geometric graph from which we removed edges of the two complete subgraphs on points of the two color classes. The only difference is that in the latter case, we can swap colors on some connected components of G. A related question was posed by Micha Perles in the DIMACS Workshop on Geometric Graph Theory in 2002. He asked what the maximum number $h(n)$ is such that if we remove arbitrary $h(n)$ edges from a complete geometric graph on an arbitrary set of n points in general position, we can still find a noncrossing Hamiltonian path. Černý et al. [6] showed that $h(n) = \Omega(\sqrt{n})$ and also that it is safe to remove the edges of an arbitrary complete graph on $\Omega(\sqrt{n})$ vertices and that this bound is asymptotically optimal. Aichholzer et al. [3] summarize history and results of this type also for graphs different from the path.

It is thus impossible to find a noncrossing Hamiltonian alternating (that is, properly colored) path (*NHAP* for short) on some bicolored point sets. Kaneko et al. [14] proved that the smallest such point set has 16 points if we allow only an even number n of points and 13 for arbitrary n.

Several sufficient conditions are known under which a bicolored point set admits an NHAP. An NHAP exists whenever the two color classes are separable by a line [1] or if one of them is composed of the points of the convex hull of S [1].

The result on sets with color classes separable by a line readily implies that any 2-colored set S with each color class of size $n/2$ admits a noncrossing alternating path (NAP) on at least $n/2$ points of S. It is an open problem if this lower bound can be improved to $n/2 + f(n)$, where $f(n)$ is unbounded (see also Sect. 9.7 of the book [5]). On the other hand, there are such 2-colored sets admitting no NAP of length more than $\approx 2n/3$ [2, 16]. This upper bound is proved for certain colorings of points in convex position. The above general lower bound $n/2$ can be slightly improved for sets in convex position [12, 16]; the best bound is currently $n/2 + \Omega(\sqrt{n})$ by Hajnal and Mészáros [12].

The main result of this chapter is that some point sets contain an NHAP for any equitable 2-coloring of their points. We call such point sets 2-*color universal* for a path.

See the survey of Kaneko and Kano [13] for more results on embedding graphs on bicolored point sets. One of the results mentioned in the survey is the possibility to embed some graphs with a fixed 2-coloring on an arbitrary compatibly 2-colored point set. Let G be a forest of stars with centers colored black and leaves white, and let S be a 2-colored point set. If we map the centers of stars arbitrarily and then we map the leaves so that the sum of lengths of edges is minimized, then no two edges cross.

Previously, a different notion of universality was considered in the context of embedding colored graphs on colored point sets. A k-colored set S of n points is *universal* for a class $\mathcal{G}$ of graphs if every (not necessarily proper) coloring of vertices of $G \in \mathcal{G}$ on n vertices admits an embedding on S. Brandes et al. [4] find, for example, universal k-colored point sets for the class of caterpillars for every $k \leq 3$.

1.2 Our Results

A *convex* or a *concave chain* is a finite set of points in the plane lying on the graph of a strictly convex or a strictly concave function, respectively. A *double-chain* (C_1, C_2) consists of a convex chain C_1 and a concave chain C_2 such that each point of C_2 lies strictly below every line determined by C_1, and similarly, each point of C_1 lies strictly above every line determined by C_2 (see Fig. 1). Double-chains were first considered in [8].

The size of a chain C_i is the number $|C_i|$ of its points. Note that we allow different sizes of the chains C_1 and C_2. If the sizes $|C_1|$, $|C_2|$ of the chains differ by at most 1, we say that the double-chain is *balanced*.

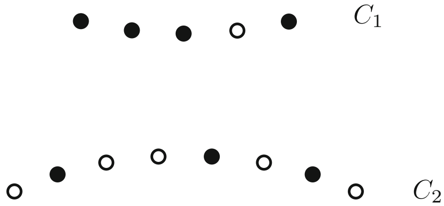


Fig. 1 An equitable 2-coloring of a double-chain (C_1, C_2)

We consider only 2-colorings and we use *black* and *white* as the colors. A 2-coloring of a set S of n points in the plane is *compatible* with a 2-coloring of a graph G on n vertices if the number of black points of S is the same as the number of black vertices in G. This implies the equality of numbers of white points and white vertices as well.

A graph G with 2-colored vertices is *embeddable* on a 2-colored point set S if the vertices of G can be mapped to the points of S so that the colors match and no two edges cross if they are drawn as straight-line segments.

A set S of points is 2-*color universal* for a bipartite graph G if, for every proper 2-coloring of G and for every compatible 2-coloring of S, G is embeddable on S.

If the properly colored path on n vertices, P_n, can be embedded on a 2-colored set S of n points, then we say that S has an NHAP (noncrossing Hamiltonian alternating path).

A 2-coloring of a set S of n points is *equitable* if it is compatible with some proper 2-coloring of P_n, that is, if the sizes of the two color classes differ by at most 1.

Section 2 contains the proof of the following theorem.

Theorem 1. *Let (C_1, C_2) be a double-chain satisfying $|C_i| \geq 1/5(|C_1| + |C_2|)$ for $i = 1, 2$. Then (C_1, C_2) is 2-color universal for P_n, where $n = |C_1| + |C_2|$. Moreover, an NHAP on an equitably colored (C_1, C_2) can be found in linear time.*

Note that this doesn't always hold if we don't require the coloring of the path to be proper. For example, if we color first two vertices of P_4 black and the other two white, then it cannot be embedded on the double-chain with both chains of size 2 if the two black points are the top left one and the bottom right one.

In Sect. 3, we show that double-chains with highly unbalanced sizes of chains do not admit an NHAP for some equitable 2-colorings.

Theorem 2. *Let (C_1, C_2) be a double-chain, and let C_1 be periodic with the following period of length 16: 2 black, 4 white, 6 black, and 4 white points. If $|C_1| \geq 28(|C_2| + 1)$, then (C_1, C_2) has no NHAP.*

An *equitable coloring* of a graph is a coloring where the sizes of any two color classes differ by at most 1. A bipartite graph is *equitable* if it admits a proper equitable 2-coloring.

Section 4 mainly studies other graphs for which the balanced double-chain is 2-color universal.

Theorem 3. *The balanced double-chain is* 2*-color universal for equitable caterpillars with at least as many leaves as nonleaf vertices.*

If a forest of stars is 2*-colored equitably and properly, then it can be embedded on every compatibly* 2*-colored balanced double-chain.*

If the balanced double-chain is 2*-color universal for an equitable bipartite graph* G*, then* G *is a forest of caterpillars.*

We also present examples of equitable bipartite planar graphs for which no set of points is 2-color universal.

2 Proof of Theorem 1

The main idea of our proof is to cover the chains C_i by a special type of pairwise non-crossing paths, so called hedgehogs, and then to connect these hedgehogs into an NHAP by adding some edges between C_1 and C_2.

2.1 Notation Used in the Proof

For $i = 1,2$, let b_i be the number of black points of C_i and let $w_i := |C_i| - b_i$ denote the number of white points of C_i.

Since the coloring is equitable, we may assume that $b_1 \geq w_1$ and $w_2 \geq b_2$. Then black is *the major color of* C_1 and *the minor color of* C_2, and white is *the major color of* C_2 and *the minor color of* C_1. Points in the major color, i.e., black points on C_1 and white points on C_2, are called *major points*. Points in the minor color are called *minor points*.

Points on each C_i are linearly ordered according to the x-coordinate. An *interval* of C_i is a sequence of consecutive points of C_i. An *inner point* of an interval I is any point of I that is neither the leftmost nor the rightmost point of I.

A *body* D is a nonempty interval of a chain C_i $(i = 1,2)$ such that all inner points of D are major. If the leftmost point of D is minor, then we call it a *head* of D. Otherwise, D has no head. If the rightmost point of D is minor, then we call it a *tail* of D. Otherwise, D has no tail. If a body consists of just one minor point, this point is both the head and the tail.

Bodies are of the following four types. A 00*-body* is a body with no head and no tail. A 11*-body* is a body with both head and tail. The bodies of the remaining two types have exactly one endpoint major and the other one minor. We will call the body a 10*-body* or a 01*-body* if the minor endpoint is a head or a tail, respectively.

Let D be a body on C_i. A *hedgehog (built on the body* $D \subseteq C_i$*)* is a noncrossing alternating path H with vertices in C_i satisfying the following three conditions: (1) H contains all points of D; (2) H contains no major points outside D; (3) the endpoints of H are the first and last points of D. A hedgehog built on an $\alpha\beta$-body

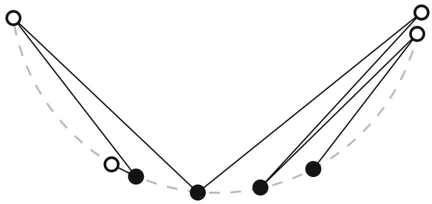


Fig. 2 A hedgehog in C_1

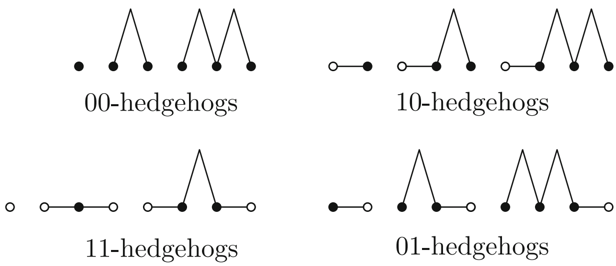


Fig. 3 Examples of hedgehogs

is an $\alpha\beta$-*hedgehog* ($\alpha, \beta = 0, 1$). See Fig. 2. If a hedgehog H is built on a body D, then D is *the body of* H and the points of H that do not lie in D are *spines*. Note that each spine is a minor point. All possible types of hedgehogs can be seen in Fig. 3 (for better lucidity, we will draw hedgehogs with bodies on a horizontal line and spines indicated only by a "peak" from now on).

On each C_i, maximal intervals containing only major points are called *runs*. Clearly, runs form a partition of major points. For $i = 1, 2$, let r_i denote the number of runs in C_i.

2.2 Proof in the Even Case

Throughout this subsection, (C_1, C_2) denotes a double-chain with $|C_1| + |C_2|$ even. Since the coloring is equitable, we have $b_1 + b_2 = w_1 + w_2$. Set

$$\Delta := b_1 - w_1 = w_2 - b_2.$$

First, we give a lemma characterizing collections of bodies on a chain C_i that are bodies of some pairwise noncrossing hedgehogs covering the whole chain C_i.

Lemma 2.1. *Let $i \in \{1, 2\}$. Let all major points of C_i be covered by a set $\mathcal{D}$ of pairwise disjoint bodies. Then the bodies of $\mathcal{D}$ are the bodies of some pairwise noncrossing hedgehogs covering the whole C_i if and only if $\Delta = d_{00} - d_{11}$, where $d_{\alpha\alpha}$ is the number of $\alpha\alpha$-bodies in $\mathcal{D}$.*

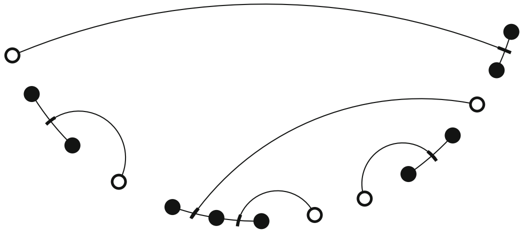


Fig. 4 A noncrossing matching of minor points and midpoints (in C_1)

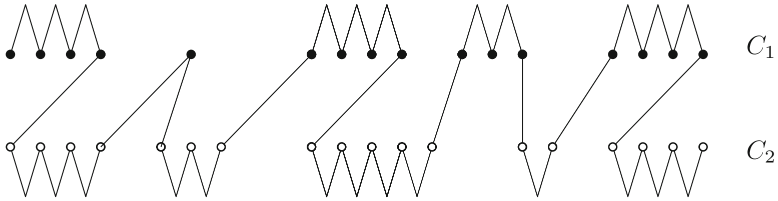


Fig. 5 00-Hedgehogs connected to an NHAP

Proof. An $\alpha\beta$-hedgehog containing t major points contains $(t-1)+\alpha+\beta$ minor points. It follows that the equality $\Delta = d_{00} - d_{11}$ is necessary for the existence of a covering of C_i by disjoint hedgehogs built on the bodies of $\mathcal{D}$.

Suppose now that $\Delta = d_{00} - d_{11}$. Let F be the set of minor points on C_i that lie in no body of $\mathcal{D}$, and let M be the set of the midpoints of straight-line segments connecting pairs of consecutive major points lying in the same body. It is easily checked that $|F| = |M|$. Clearly, $F \cup M$ is a convex or a concave chain. Now it is easy to prove that there is a noncrossing perfect matching formed by $|F| = |M|$ straight-line segments between F and M (for the proof, take any segment connecting a point of F with a neighboring point of M, remove the two points, and continue by induction); see Fig. 4.

If $f \in F$ is connected to a point $m \in M$ in the matching, then f will be a spine with edges going from it to those two major points that determined m. Obviously, these spines and edges define noncrossing hedgehogs with bodies in $\mathcal{D}$ and with all the required properties. □

The following three lemmas and their proofs show how to construct an NHAP in some special cases.

Lemma 2.2. *If* $\Delta \geq \max\{r_1, r_2\}$*, then* (C_1, C_2) *has an NHAP.*

Proof. Let $i \in \{1,2\}$. Since $r_i \leq \Delta \leq \max(b_i, w_i)$, the runs in C_i may be partitioned into Δ 00-bodies. By Lemma 2.1, these 00-bodies may be extended to pairwise noncrossing hedgehogs covering C_i. This gives us 2Δ hedgehogs on the double-chain. They may be connected into an NHAP by $2\Delta - 1$ edges between the chains in the way shown in Fig. 5. □

Lemma 2.3. *If* $r_1 = r_2$*, then* (C_1, C_2) *has an NHAP.*

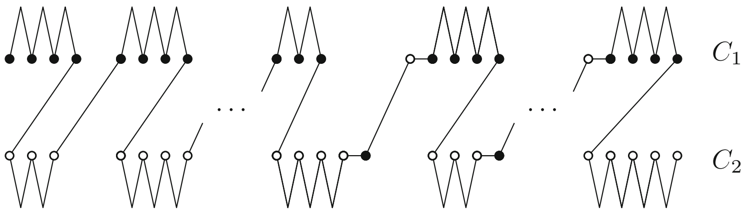


Fig. 6 An NHAP in the case $r_1 = r_2 > \Delta \geq 1$

Proof. Set $r := r_1 = r_2$. If $r \leq \Delta$, then we may apply Lemma 2.2. Thus, let $r > \Delta$.

Suppose first that $\Delta \geq 1$. We cover each run on each C_i by a single body whose type is as follows. On C_1 we take Δ 00-bodies followed by $(r-\Delta)$ 10-bodies. On C_2 we take (from left to right) $(\Delta-1)$ 00-bodies, $(r-\Delta)$ 01-bodies, and one 00-body. By Lemma 2.1, the r bodies on each C_i can be extended to hedgehogs covering C_i. Altogether we obtain $2r$ hedgehogs. They can be connected to an NHAP by $2r-1$ edges between C_1 and C_2 (see Fig. 6).

Suppose now that $\Delta = 0$. We add one auxiliary major point on each C_i as follows. On C_1, the auxiliary point extends the leftmost run on the left. On C_2, the auxiliary point extends the rightmost run on the right. This does not change the number of runs and increases Δ to 1. Thus, we may proceed as above. The NHAP obtained has the two auxiliary points on its ends. We may remove the auxiliary points from the path, obtaining an NHAP for (C_1, C_2). $\square$

A *singleton* $s \in C_i$ is an inner point of C_i $(i = 1, 2)$ such that its two neighbors on C_i are colored differently from s.

Lemma 2.4. *Suppose that C_1 has no singletons and C_2 can be covered by $r_1 - 1$ pairwise disjoint hedgehogs. Then (C_1, C_2) has an NHAP.*

Proof. For simplicity of notation, set $r := r_1$. We denote the $r-1$ hedgehogs on C_2 by $P_1, P_2, \ldots, P_{r-1}$ in the left-to-right order in which the bodies of these hedgehogs appear on C_2. For technical reasons, we enlarge the leftmost run of C_1 from the left by an auxiliary major point σ.

Our goal is to find r hedgehogs $H_1, H_2, \ldots, H_r$ on $C_1 \cup \{\sigma\}$ such that they may be connected with the hedgehogs $P_1, P_2, \ldots, P_{r-1}$ into an NHAP. For each $j = 1, \ldots, r$, the body of the hedgehog H_j will be denoted by D_j. For each $j = 1, \ldots, r$, D_j covers the jth run of $C_1 \cup \{\sigma\}$ (in the left-to-right order). We now finish the definition of the bodies D_j by specifying for each D_j if it has a head and/or a tail. The body D_1 is without a head. For $j > 1$, D_j has a head if and only if P_{j-1} has a tail. The last body D_r is without a tail and $D_j, j < r$ has a tail if and only if P_j has a head.

It follows from Lemma 2.1 that we may add or remove some minor points on $C_1 \cup \{\sigma\}$ so that $D_1, \ldots, D_r$ can then be extended to pairwise noncrossing hedgehogs $H_1, \ldots, H_r$ covering the "new" C_1. More precisely, there is a double-chain (C_1', C_2) such that $D_1, \ldots, D_r$ can be extended to pairwise noncrossing hedgehogs $H_1, \ldots, H_r$ covering C_1', where either $C_1' = C_1 \cup \{\sigma\}$ or C_1' is obtained from $C_1 \cup \{\sigma\}$ by adding

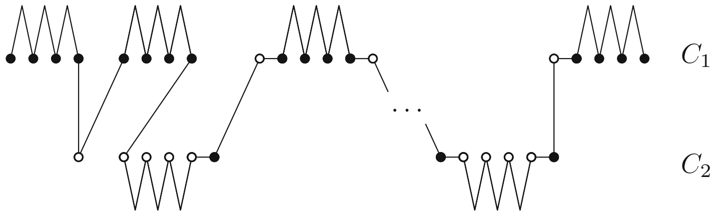


Fig. 7 An NHAP in the case of no singletons on C_1

some minor (white) points on the left of $C_1 \cup \{\sigma\}$ (say) or C_1' is obtained from $C_1 \cup \{\sigma\}$ by removing some minor (white) points lying in none of the bodies $D_1, \ldots, D_r$. Then the concatenation $H_1P_1H_2P_2 \cdots H_{r-1}P_{r-1}H_r$ shown in Fig. 7 gives an NHAP on (C_1', C_2). This NHAP starts with the point σ. Removing σ from it gives an NHAP P for the double-chain $(C_1' \setminus \{\sigma\}, C_2)$. The endpoints of P have different colors. Thus, P covers the same number of black and white points. Black points on P are the $(|C_1| + |C_2|)/2$ black points of (C_1, C_2). Thus, P covers exactly $|C_1| + |C_2|$ points. It follows that $|C_1' \setminus \{\sigma\}| = |C_1|$, and thus $C_1' \setminus \{\sigma\} = C_1$. The path P is an NHAP on the double-chain (C_1, C_2). □

The following lemma will be used to find the type of covering given in the assumption of Lemma 2.4.

Lemma 2.5. *Suppose that $|C_i| \geq k$, $r_i \leq k$, and $\Delta \leq k$ for some $i \in \{1,2\}$ and for some integer k. Then C_i can be covered by k pairwise disjoint hedgehogs.*

Proof. The idea of the proof is to start with the set $\mathcal{D}$ of $|C_i|$ bodies, each of them being a single point, and then gradually decrease the number of bodies in $\mathcal{D}$ by joining some of the bodies together. We see that $\Delta = d_{00} - d_{11}$, where $d_{\alpha\alpha}$ is the number of $\alpha\alpha$-bodies in $\mathcal{D}$. If we join two neighboring 00-bodies to one 00-body and withdraw a single-point 11-body from $\mathcal{D}$ (to let the minor point become a spine) at the same time, the difference between the number of 00-bodies and the number of 11-bodies remains the same and $|\mathcal{D}|$ decreases by 2. We can reduce $|\mathcal{D}|$ by 1 while preserving the difference $d_{00} - d_{11}$ by joining a 00-body with a neighboring single-point 11-body into a 01- or a 10-body. Similarly, we can join a 01- or 10-body with a neighboring (from the proper side) single-point 11-body into a new 11-body to decrease $|\mathcal{D}|$ by 1 as well. When we are joining two 00-bodies, we choose the single-point 11-body to remove in such a way to keep as many single-point 11-bodies adjacent to 00-bodies as possible. This guarantees that we can use up to r_i of them for heads and tails.

We start with joining neighboring 00-bodies, and we do this as long as $|\mathcal{D}| > k+1$ and $d_{00} > r_i$. Note that by the assumption $\Delta \leq k$, we will have enough single-point 11-bodies to do that. When we end, one of the following conditions holds: $|\mathcal{D}| = k$, $|\mathcal{D}| = k+1$, or $d_{00} = r_i$. In the first case, we are done. If $|\mathcal{D}| = k+1$, we just add one head or one tail (we can do this since $d_{00} + d_{11} = |\mathcal{D}| = k+1 \geq d_{00} - d_{11} + 1$, which implies $d_{11} > 0$). If $d_{00} = r_i$, then each run is covered by just one 00-body.

We need to add $|\mathcal{D}|-k$ heads and tails. We have enough single-point 11-bodies to do that since $d_{11}=|\mathcal{D}|-d_{00}=|\mathcal{D}|-r_i \geq |\mathcal{D}|-k$. On the other hand, $r_i-d_{11}=\Delta \geq 0$, so the number of heads and tails needed is at most r_i. Therefore, all the single-point 11-bodies are adjacent to 00-bodies, and we can use them to form heads and tails.

In all cases we get a set $\mathcal{D}$ of k bodies. Now we can apply Lemma 2.1 to obtain k pairwise disjoint hedgehogs covering C_i. □

By a *contraction* we mean removing a singleton with both its neighbors and putting a point of the color of its neighbors in its place instead. It is easy to verify that if there is an NHAP in the new double-chain obtained by this contraction, it can be expanded to an NHAP in the original double-chain.

Now we can prove our main theorem in the even case.

proof of theorem 1.1 (even case). Without loss of generality, we may assume that $r_1 \geq r_2$. In the case $r_1 = r_2$, we get an NHAP by Lemma 2.3. In the case $\Delta \geq r_1$, we get an NHAP by Lemma 2.2. Therefore, the only case left is $r_1 > r_2$, $r_1 > \Delta$.

If there is a singleton on C_1, we make a contraction of it. The contraction decrease r_1 by 1 and both r_2 and Δ remain unchanged. If now $r_1 = r_2$ or $r_1 = \Delta$, we again get an NHAP; otherwise, we keep making contractions until one of the previous cases appears or there are no more singletons to contract.

If there is no more singleton to contract on C_1 and still $r_1 > r_2$ and $r_1 > \Delta$, we try to cover C_2 by $r_1 - 1$ pairwise disjoint paths. Before the contractions, $|C_2| \geq |C_1|/4$ did hold, and by the contractions we could just decrease $|C_1|$; therefore, it still holds.

All the maximal intervals on the chain C_1 (with possible exception of the first and the last one) have now length at least 2, which implies that $r_1 \leq |C_1|/4+1$. Hence, $|C_2| \geq |C_1|/4 \geq r_1 - 1$, so we can create $r_1 - 1$ pairwise disjoint hedgehogs covering C_2 using Lemma 2.5. Then we apply Lemma 2.4 and expand the NHAP obtained by Lemma 2.4 to an NHAP on the original double-chain.

There is a straightforward linear-time algorithm for finding an NHAP on (C_1, C_2) based on the above proof. □

2.3 Proof in the Odd Case

In this subsection, we prove Theorem 1 for the case when $|C_1|+|C_2|$ is odd. We set $\Delta = w_2 - b_2$ and proceed similarly as in the even case. In several places in the proof, we will add one auxiliary point ω to get the even case (its color will be chosen to equalize the numbers of black and white points). We will be able to apply one of the Lemmas 2.2–2.4 to obtain an NHAP. The point ω will be at some end of the NHAP, and by removing ω, we obtain an NHAP for (C_1, C_2).

Without loss of generality, we may assume that $r_1 \geq r_2$. In the case $r_1 = r_2$, we add an auxiliary major point ω, which is placed either as the left neighbor of the leftmost major point on C_1 or as the right neighbor of the rightmost major point on C_2. Then we get an NHAP by Lemma 2.3, and the removal of ω gives us an NHAP for (C_1, C_2).

In the case $\Delta \geq r_1$, we add an auxiliary point ω to the same place and we get an NHAP by Lemma 2.2. Again, the removal of ω gives us an NHAP for (C_1, C_2).

Now, the only case left is $r_1 > r_2$, $r_1 > \Delta$. If there are any singletons on C_1, we make the contractions exactly the same way as in the proof of the even case. If Lemma 2.2 or 2.3 needs to be applied, we again add an auxiliary point ω and proceed as above.

If there is no more singleton to contract on C_1 and still $r_1 > r_2$ and $r_1 > \Delta$, we have $|C_2| \geq |C_1|/4 \geq r_1 - 1$ as in the proof of the even case and we can use Lemma 2.5 to get $r_1 - 1$ pairwise disjoint hedgehogs covering C_2. Now we need to consider two cases: (1) If $b_1 + b_2 > w_1 + w_2$, then we find an NHAP for (C_1, C_2) in the same way as in the proof of Lemma 2.4, except we do not add the auxiliary point σ. (2) If $b_1 + b_2 < w_1 + w_2$, we add an auxiliary point ω as the right neighbor of the rightmost major point on C_1. The number r_1 didn't change, so Lemma 2.4 gives us an NHAP. Again, the removal of ω gives us an NHAP for (C_1, C_2).

There is a straightforward linear-time algorithm for finding an NHAP on (C_1, C_2) based on the above proof. □

3 Unbalanced Double-Chains with No NHAP

In this section, we prove Theorem 2. Let (C_1, C_2) be a double-chain whose points are colored by an equitable 2-coloring, and let C_1 be periodic with the following period: 2 black, 4 white, 6 black, and 4 white points. Let $|C_1| \geq 28(|C_2| + 1)$. We want to show that (C_1, C_2) has no NHAP.

Suppose on the contrary that (C_1, C_2) has an NHAP. Let $P_1, P_2, \ldots, P_t$ denote the maximal subpaths of the NHAP containing only points of C_1. Since between every two consecutive paths P_i, P_j in the NHAP there is at least one point of C_2, we have $t \leq |C_2| + 1$. In the following we think of C_1 as a cyclic sequence of points on the circle. Note that we get more intervals in this way. Theorem 2 now directly follows from the following theorem, which will be proven in the rest of this section.

Theorem 3.1. *Let C_1 be a set of points on a circle periodically 2-colored with the following period of length 16: 2 black, 4 white, 6 black, and 4 white points. Suppose that all points of C_1 are covered by a set of t noncrossing alternating and pairwise disjoint paths $P_1, P_2, \ldots, P_t$. Then $t > |C_1|/28$.*

Each maximal interval spanned by a path P_i on the circle is called a *base*. Let $b(P_i)$ denote the number of bases of P_i. A path with one base only is called a *leaf*. We consider the following special types of edges in the paths. *Long edges* connect points that belong to different bases. *Short edges* connect consecutive points on C_1. Note that short edges cannot be adjacent to each other. A maximal subpath of a path P_i spanning two subintervals of two different bases and consisting of long edges only is called a *zigzag*. A path is *separated* if all of its edges can be crossed by a line. Note that each zigzag is a separated path. A maximal separated subpath of P_i that contains an endpoint of P_i and spans one interval only is a *rainbow*. We find all the zigzags and rainbows in each P_i, $i = 1,2,\ldots,t$. Note that two zigzags, or a zigzag and a rainbow, are either disjoint or share an endpoint. A *branch* is a maximal subpath of P_i that spans two intervals and is induced by a union of zigzags.

For each path P_i that is not a leaf, construct the following graph G_i. The vertices of G_i are the bases of P_i. We add an edge between two vertices for each branch that connects the corresponding bases. If G_i has a cycle (including the case of a "2-cycle"), then one of the corresponding branches consists of a single edge that lies on the convex hull of P_i. We delete such an edge from P_i and no longer call it a branch. By deleting a corresponding edge from each cycle of G_i, we obtain a graph G'_i, which is a spanning tree of G_i. The *branch graph* G' is a union of all graphs G'_i.

Let $\mathcal{L}$ denote the set of leaves and $\mathcal{B}$ the set of branches. Let $\mathcal{P} = \{P_1, P_2, \ldots, P_t\}$.

Observation 3.2. *The branch graph G' is a forest with components G'_i. Therefore,*

$$|\mathcal{B}| = \sum_{i,P_i \notin \mathcal{L}} (b(P_i) - 1). \qquad \square$$

The branches and rainbows in P_i do not necessarily cover all the points of P_i. Each point that is not covered is adjacent to a deleted long edge and to a short edge that connects this point to a branch or a rainbow. It follows that between two consecutive branches (and between a rainbow and the nearest branch), there are at most two uncovered points, which are endpoints of a common deleted edge. By an easy case analysis, it can be shown that this upper bound can be achieved only if one of the nearest branches consists of a single zigzag.

In the rest of the chapter, a *run* will be a maximal monochromatic interval of any color. In the following, we will count the runs that are spanned by the paths P_i. The *weight* of a path P, $w(P)$, is the number of runs spanned by P. If P spans a whole run, it adds one unit to $w(P)$. If P partially spans a run, it adds half a unit to $w(P)$.

Observation 3.3. *The weight of a zigzag or a rainbow is at most* 1.5. *A branch consists of at most two zigzags; hence, it weighs at most three units.* $\square$

Lemma 3.4. *A path P_i that is not a leaf weighs at most* $3.5k + 3.5$ *units, where k is the number of branches in P_i.*

Proof. According to the above discussion, for each pair of uncovered points that are adjacent on P_i, we can join one of them to the adjacent branch consisting of a single zigzag. To each such branch we join at most two uncovered points; hence, its weight increases by at most one unit to at most 2.5 units. The number of the remaining uncovered points is at most $k+1$. Therefore, $w(P_i) \leq 3k + 3 + 0.5 \cdot (k+1) = 3.5k + 3.5$. $\square$

Lemma 3.5. *A leaf weighs at most* 3.5 *units.*

Proof. Let L be a leaf spanning at least two points. Consider the interval spanned by L. Cut this interval out of C_1 and glue its endpoints together to form as circle. Take a line l that crosses the first and last edges of L. Note that the line l doesn't separate any of the runs. Exactly one of the arcs determined by l contains the gluing point γ.

Each of the ending edges of L belongs to a rainbow, all of whose edges cross l. It follows that if L has only one rainbow, then this rainbow covers the whole leaf L and $w(L) \leq 1.5$. Otherwise, L has exactly two rainbows, R_1 and R_2. We show that R_1 and R_2 cover all edges of L that cross the line l. Suppose there is an edge s in L that crosses l and does not belong to any of the rainbows R_1, R_2. Then one of these rainbows, say R_1, is separated from γ by s. Then the edge of L that is the second-nearest to R_1 also has the same property as the edge s. This would imply that R_1 spans two whole runs, a contradiction. It follows that all the edges of L that are not covered by the rainbows are consecutive and connect adjacent points on the circle. There are at most three such edges; at most one connecting the points adjacent to γ, the rest of them being short on C_1. But this upper bound of three cannot be achieved since it would force both rainbows to span two whole runs. Therefore, there are at most two edges and hence at most one point in L uncovered by the rainbows. The lemma follows. □

Lemma 3.6. $|\mathcal{L}| \geq \sum_{i,P_i \notin \mathcal{L}} (b(P_i) - 2) + 2.$

Proof. The number of runs in C_1 is at least 4. By Lemma 3.5, if all the paths P_i are leaves, then at least 2 of them are needed to cover C_1, and the lemma follows.

If not all the paths are leaves, we order the paths so that all the leaves come at the end of the ordering. The path P_1 spans $b(P_1)$ bases. Shrink these bases to points. These points divide the circle into $b(P_1)$ arcs, each of which contains at least one leaf. If P_2 is not a leaf, then continue. The path P_2 spans $b(P_2)$ intervals on one of the previous arcs. Shrink them to points. These points divide the arc into $b(P_2)+1$ subarcs. At least $b(P_2) - 1$ of them contain leaves. This increased the number of leaves by at least $b(P_2) - 2$. The case of P_i, $i > 2$, is similar to P_2. The lemma follows by induction. □

Corollary 3.7. $|\mathcal{B}| \leq |\mathcal{P}| - 2.$

Proof. Combining Lemma 3.6 and Observation 3.2, we get the following:

$$|\mathcal{B}| = \sum_{i,P_i \notin \mathcal{L}} (b(P_i) - 1) = \sum_{i,P_i \notin \mathcal{L}} (b(P_i) - 2) + |\mathcal{P}| - |\mathcal{L}| + 2 - 2 \leq |\mathcal{P}| - 2. \qquad \square$$

Now we are in position to finish the proof of Theorem 3.1. If the whole C_1 is covered by the paths P_i, then $\sum_{i=1}^{t} w(P_i) \geq |C_1|/4$. Therefore,

$$|C_1| \leq 4 \cdot (3.5|\mathcal{B}| + 3.5(|\mathcal{P}| - |\mathcal{L}|) + 3.5|\mathcal{L}|) < 4 \cdot 7|\mathcal{P}| = 28|\mathcal{P}|.$$

4 Embedding Equitable Bipartite Graphs

4.1 Embedding on Balanced Double-Chains

We already know that the balanced double-chain is 2-color universal for the path P_n. In this subsection, we further study the class of graphs for which the balanced double-chain is 2-color universal. The three lemmas of this subsection prove the three claims of Theorem 3.

Lemma 4.1. *If the balanced double-chain is* 2*-color universal for an equitable bipartite graph G, then G is a forest of caterpillars.*

Proof. Let $K_{1,3}^+$ be the 3-star with subdivided edges (see Fig. 8). A connected graph is a caterpillar if and only if it contains no cycle and no $K_{1,3}^+$ as a subgraph.

We will color all points of one chain white and points of the other chain black so that the resulting coloring is compatible with the 2-coloring of G. We assume for contradiction that G can be embedded on it and that it contains either a cycle or $K_{1,3}^+$.

As the double-chain has monochromatic chains, all the edges connect the two chains. Because the embedding has no edge crossings, we can consider the leftmost edge of the cycle and let u and v be its end vertices. Then the two edges of the cycle incident to exactly one of u and v cross.

We now assume that $K_{1,3}^+$ can be embedded on the double-chain and let the color of its root vertex be white. Let ω_1 be the white point where the root of $K_{1,3}^+$ is mapped, and let β_1, β_2, β_3 be (from left to right) the three black points where the middle vertices are mapped. Then β_2 is connected by an edge to some white leaf vertex of $K_{1,3}^+$, but this edge is crossed either by the segment $\omega_1\beta_1$ or by $\omega_1\beta_3$. See Fig. 8. □

The *central path* in a caterpillar is the set of nonleaf vertices.

Lemma 4.2. *If an equitable bipartite graph G on n vertices is a caterpillar with at most* $\lfloor n/2 \rfloor$ *vertices on the central path, then the balanced double-chain is* 2*-color universal for G.*

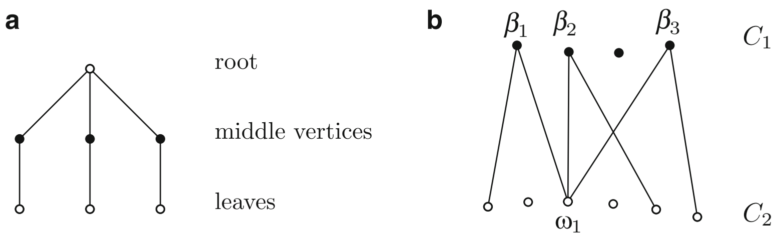


Fig. 8 (**a**) $K_{1,3}^+$ and (**b**) impossibility of its embedding on the double-chain with monochromatic chains

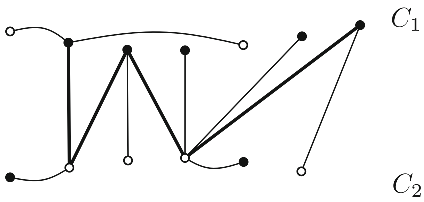


Fig. 9 Embedding a caterpillar on a balanced double-chain; the *bold edges* form the central path

Proof. Let b_i be the number of black points on the chain C_i and w_i the number of white points on C_i. Since the coloring is equitable, we may assume that $b_1 \geq w_1$ and $w_2 \geq b_2$.

Observe that the number of minor points on each chain is at most the number of leaves of G of that color. Let G' be the graph with b_1 black and w_1 white vertices obtained from G by removing some b_2 black and w_2 white leaves.

In the first phase, we embed G' on the set of major points of the two chains. We take the vertices of the central path of G' starting from one of its ends. A vertex v of the central path is placed on the leftmost unused major point on the chain where the color of v is major. The leaves of v in G' are then successively placed on the leftmost unused major points of the other chain.

In the second phase, we map all the leaves removed in the first phase on minor points. In the same greedy way as in the proof of Lemma 2.1, we keep connecting the closest pair of an unused white point of C_1 and a black point of the central path that still misses at least one leaf. The same is done on C_2.

This guarantees that no crossing appears and that every vertex is mapped to some point. See Fig. 9. □

Lemma 4.3. *If a forest of stars G is 2-colored equitably and properly, then G can be embedded on every compatibly 2-colored balanced double-chain.*

Proof. We take some fixed proper equitable 2-coloring of G.

We show that by adding edges to G, we are able to create a properly 2-colored caterpillar on the set of all vertices of G and with at most $\lfloor n/2 \rfloor$ nonleaf vertices. By Lemma 4.2, this caterpillar can be embedded on every compatibly 2-colored balanced double-chain, and thus G can be embedded.

The cases when $n \leq 3$ and when G has no edges are trivial.

For every $i \geq 3$, let k_i (h_i) be the number of stars on i vertices and with black (white) central vertex. In case of 2-vertex components of G, we cannot distinguish the central vertex, and we let n_2 be their number. We also let n_1 be the total number of 1-vertex components of G, as it is not necessary for the proof to count black and white ones separately.

We assume without loss of generality that at least half of the stars on at least three vertices have a black central vertex. We start connecting the central vertices of stars on at least three vertices into an alternating path starting with a black vertex. At some point we run out of stars with a white central vertex. We then use the stars on two vertices as stars with a white central vertex. If we run out of stars with a black

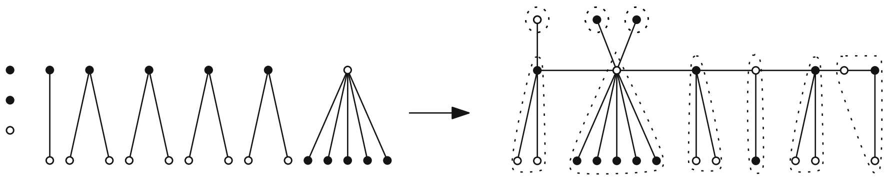


Fig. 10 Connecting stars to form a caterpillar

central vertex, we use every second star on two vertices as a star with a black central vertex. Otherwise, we run out of stars on two vertices. Then we start connecting each of the remaining stars on at least three vertices by an edge between one of its white leaves and the last black vertex on the path.

The resulting graph is composed of a connected graph T and all 1-vertex components of G. The graph T is a properly colored caterpillar and the created path is its central path P. Since G has some edges, P is not empty.

If P contains only one vertex v, we pick one of its leaves, u, and connect every 1-vertex component of G either to u or to v, depending on its color. The central path then has 2 vertices, which is at most $\lfloor n/2 \rfloor$.

If P has at least two vertices, we connect every 1-vertex component of G to a vertex of the other color on the central path. See Fig. 10.

It remains to show that the central path is not too long. The total number of vertices is

$$n = n_1 + 2n_2 + \sum_{i=3}^{n} i(k_i + h_i). \tag{1}$$

If $\sum_{i=3}^{n} k_i \leq n_2 + \sum_{i=3}^{n} h_i$, then every vertex of the central path has at least one leaf, and thus the caterpillar has at most $\lfloor n/2 \rfloor$ nonleaf vertices.

Otherwise, the central path starts and ends with a black vertex, and the black vertices of the central path are exactly the black centers of stars on at least three vertices. The central path thus has $2\sum_{i=3}^{n} k_i - 1$ vertices.

Because the 2-coloring of G is equitable, the number of black vertices of G is at least $\lfloor n/2 \rfloor$, and thus

$$\sum_{i=3}^{n} k_i + \sum_{i=3}^{n} (i-1)h_i + n_2 + n_1 \geq \left\lfloor \frac{n}{2} \right\rfloor.$$

At most $\lfloor n/2 \rfloor + 1$ vertices of G are white, which leads to

$$\left\lfloor \frac{n}{2} \right\rfloor + 1 \geq \sum_{i=3}^{n} h_i + \sum_{i=3}^{n} (i-1)k_i + n_2.$$

Putting the two inequalities together gives us

$$\sum_{i=3}^{n} (i-2)h_i + n_1 \geq \sum_{i=3}^{n} (i-2)k_i - 1. \tag{2}$$

Fig. 11 Colorings of double-chains not admitting $K^+_{1,4}$

The number of vertices of the central path is at most $\lfloor n/2 \rfloor$, because

$$\begin{aligned} 2\left(2\sum_{i=3}^{n} k_i - 1\right) &\leq \sum_{i=3}^{n} 2(i-1)k_i - 2 \\ &< \sum_{i=3}^{n} ik_i + \sum_{i=3}^{n}(i-2)k_i - 1 \\ &\leq \sum_{i=3}^{n} ik_i + \sum_{i=3}^{n}(i-2)h_i + n_1 \qquad \text{by (2)} \\ &\leq \sum_{i=3}^{n} i(k_i + h_i) + n_1 \\ &\leq n \quad \text{by (1)}. \end{aligned}$$

□

4.2 Open Problems

It seems plausible that the balanced double-chain is 2-color universal for all equitable forests of caterpillars.

Conjecture 1. The balanced double-chain is 2-color universal for an equitable bipartite graph G if and only if G is a forest of caterpillars.

Some graphs for which the balanced double-chain is not 2-color universal have a different 2-color universal set. For example, the balanced double-chain is not 2-color universal for $K^+_{1,3}$ by Lemma 4.1, but it is easy to verify that the double-chain with one chain composed of only one vertex is.

There even exist graphs with a 2-color universal set of points, but no double-chain is 2-color universal for them. Consider the properly colored $K^+_{1,4}$ with a black central vertex. It is not embeddable on double-chains colored as in Fig. 11. But a modification of the double-chain in Fig. 12 is 2-color universal for $K^+_{1,4}$.

Some equitable bipartite planar graphs have no 2-color universal set of points.

Claim 4.4. If G is a bipartite planar quadrangulation on at least five vertices, then G has no 2-color universal set of points.

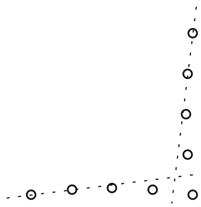


Fig. 12 A 2-color universal point set for $K_{1,4}^{+}$

Proof. Because the bipartite graph G has no 3-cycle, each of its faces has at least four vertices. Then, by Euler's formula, every planar drawing of G is a quadrangulation.

Take a set S of points and let $H(S)$ be the set of points of S on its convex hull. In a straight-line planar drawing of a graph on a set S of points, the points of $H(S)$ all lie on the outer face of the drawing. Thus, G can only be drawn on S if $3 \le |H(S)| \le 4$. In a proper coloring of G on at least five vertices, one color class contains at least three vertices. If we color three points of $H(S)$ by this color and the rest arbitrarily, G cannot be embedded, because no face in a drawing of G can contain three vertices of one color. □

The results of this chapter solve only a few particular cases of the following general question.

Question. Which planar bipartite graphs have a 2-color universal set of points?

Acknowledgements We thank Jakub Černý for his active participation in the earlier stages of our discussions.

Research was supported by the Center of Excellence – Institute for Theoretical Computer Science, Prague (project P202/12/G061 of GA CR) and by Grant SVV-2012-265313 (Discrete Models and Algorithms). Viola Mészáros was also partially supported by ESF EuroGiga project ComPose (IP03), by OTKA Grant K76099 and by OTKA Grant 102029. Josef Cibulka, Jan Kynčl, and Pavel Valtr were also supported by the Grant Agency of Charles University, GAUK 52410.

Part of the research was conducted during the Special Semester on Discrete and Computational Geometry at École Polytechnique Fédérale de Lausanne, organized and supported by the CIB (Centre Interfacultaire Bernoulli) and the SNSF (Swiss National Science Foundation).

A preliminary version appeared in proceedings of Graph Drawing 2008 [7].

References

1. M. Abellanas, J. García, G. Hernandez, M. Noy, P. Ramos, Bipartite embeddings of trees in the plane. Discr. Appl. Math. **93**, 141–148 (1999)
2. M. Abellanas, J. García, F. Hurtado, J. Tejel, C. Alternantes, in *Proceedings of X Encuentros de Geometría Computacional*, Sevilla, June 2003, pp. 7–12 (in Spanish). English version available on Ferran Hurtado's web page http://www-ma2.upc.es/~hurtado/mypapers.html
3. O. Aichholzer, S. Cabello, R. Fabila-Monroy, D. Flores-Peñaloza, T. Hackl, C. Huemer, F. Hurtado, D.R. Wood, Edge-removal and non-crossing configurations in geometric graphs. Discr. Math. Theor. Comput. Sci. **12**(1), 75–86 (2010)

4. U. Brandes, C. Erten, A. Estrella-Balderrama, J. Fowler, F. Frati, M. Geyer, C. Gutwenger, Seok-Hee Hong, M. Kaufmann, S. Kobourov, G. Liotta, P. Mutzel, A. Symvonis, Colored simultaneous geometric embeddings and universal pointsets. Algorithmica, **60**(3), 569–592 (2011). Available online at http://dx.doi.org/10.1007/s00453-010-9433-x
5. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, New York, 2005)
6. J. Černý, Z. Dvořák, V. Jelínek, J. Kára, Noncrossing Hamiltonian paths in geometric graphs. Discr. Appl. Math. **155**(9), 1096–1105 (2007)
7. J. Cibulka, J. Kynčl, V. Mészáros, R. Stolař, P. Valtr, Hamiltonian alternating paths on bicolored double-chains. in: *Graph Drawing 2008*, ed. by I.G. Tollis, M. Patrignani. Lecture Notes in Computer Science, vol. 5417 (Springer, New York, 2009), pp. 181–192
8. A. García, M. Noy, J. Tejel, Lower bounds on the number of crossing-free subgraphs of K_N. Comput. Geom. **16**, 211–221 (2000)
9. E. Di Giacomo, G. Liotta, F. Trotta, On embedding a graph on two sets of points. Int. J. Found. Comput. Sci. **17**(5), 1071–1094 (2006)
10. E. Di Giacomo, G. Liotta, F. Trotta, How to embed a path onto two sets of points. in *Graph Drawing 2005*, ed. by P. Healy, N. Nikolov. Lecture Notes in Computer Science, vol. 3843 (Springer, New York, 2006), pp. 111–116
11. P. Gritzmann, B. Mohar, J. Pach, R. Pollack, Embedding a planar triangulation with vertices at specified points. Am. Math. Mon. **98**, 165–166 (1991) (Solution to problem E3341)
12. P. Hajnal, V. Mészáros, Note on noncrossing path in colored convex sets. Discr. Math. Theor. Comput. Sci., accepted
13. A. Kaneko, M. Kano, Discrete geometry on red and blue points in the plane—a survey. in: *Discrete and Computational Geometry, The Goodman-Pollack Festschrift*, ed. by B. Aronov et al. Algorithms Combinatorial, vol. 25 (Springer, Berlin, 2003), pp. 551–570
14. A. Kaneko, M. Kano, K. Suzuki, Path coverings of two sets of points in the plane. in *Towards a Theory of Geometric Graphs, Contemporary Mathematics*. AMS, Providence, RI, vol. 342, ed. by J. Pach (2004), pp. 99–111
15. M. Kaufmann, R. Wiese, Embedding vertices at points: Few bends suffice for planar graphs. J. Graph Algorithm. Appl. **6**(1), 115–129 (2002)
16. J. Kynčl, J. Pach, G. Tóth, Long alternating paths in bicolored point sets. Discr. Math. **308**(19), 4315–4321 (2008)
17. J. Pach, R. Wenger, Embedding planar graphs at fixed vertex locations. Graphs Comb. **17**(4), 717–728 (2001)

Drawing Trees, Outerplanar Graphs, Series-Parallel Graphs, and Planar Graphs in a Small Area

Giuseppe Di Battista and Fabrizio Frati

Abstract In this chapter, we survey algorithms and bounds for constructing planar drawings of graphs in a small area.

1 Introduction

It is typical of computer science to classify problems according to the amount of resources that are needed to solve them. Hence, problems are usually classified according to the amount of time or the amount of memory that a specific model of computation requires for their solution.

This epistemological need of classifying problems finds, in the graph drawing field, a very original interpretation. A graph drawing problem can be broadly described as follows: Given a graph of a certain family and a drawing convention (e.g., all edges should be straight-line segments), draw the graph optimizing some specific features. Among those features, a fundamental one is the amount of geometric space that the drawing spans, and a natural question is: What amount of space s required to draw a planar graph, or a tree, or a bipartite graph? Hence, besides classifying problems according to the above classical coordinates, graph drawing classifies problems according to the amount of geometric space that a drawing that solves that problem requires.

Of course, such a space requirement can be influenced by the class of graphs (one can expect that the area required to draw an n-vertex tree is less than that

G. Di Battista (✉)
Dipartimento di Informatica e Automazione - Roma Tre University, Rome, Italy
e-mail: gdb@dia.uniroma3.it

F. Frati
Chair of Combinatorial Geometry - École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: frati@dia.uniroma3.it

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_9,

required to draw an n-vertex general planar graph) and by the drawing convention (straight-line drawings look more constrained than drawings where edges can be polygonal lines).

The attempt of classifying graph drawing problems with respect to the space required spurred, over the last 50 years, a large body of research. On the one hand, techniques have been devised to compute geometric lower bounds that are completely original and do not find counterparts in the techniques adopted in computer science to find time or memory lower bounds. On the other hand, the uninterrupted upper bound hunting has produced several elegant algorithmic techniques.

In this chapter, we survey the state of the art on such algorithmic and lower-bound techniques for several families of planar graphs. Indeed, drawing planar graphs without crossings is probably the most classical graph drawing topic, and many researchers have given fundamental contributions to the planar drawings of trees, outerplanar graphs, series-parallel graphs, and so forth.

We survey the state of the art, focusing on the impact of the most popular drawing conventions on the geometric space requirements. In Sect. 3, we discuss straight-line drawings. In Sect. 4, we analyze drawings where edges can be polygonal lines. In Sect. 5, we describe upward drawings, i.e., drawings of directed acyclic graphs where edges follow a common vertical direction. In Sect. 6, we describe convex drawings, where the faces of a planar drawing are constrained to be convex polygons. Proximity drawings, where vertices and edges should enforce some proximity constraints, are discussed in Sect. 7. Section 8 is devoted to drawings of clustered graphs.

We devote special attention to put in evidence those that we consider the main open problems of the field.

2 Preliminaries

In this section, we present preliminaries and definitions. For more about graph drawing, see [40, 91].

2.1 Planar Drawings, Planar Embeddings, and Planar Graphs

All the graphs that we consider are *simple*; i.e., they contain no multiple edges and loops. A *drawing* of a graph $G(V,E)$ is a mapping of each vertex of V to a point in the plane and of each edge of E to a Jordan curve connecting its endpoints. A drawing is *planar* when no two edges intersect except, possibly, at common endpoints. A *planar graph* is a graph admitting a planar drawing.

A planar drawing of a graph determines a circular ordering of the edges incident to each vertex. Two drawings of the same graph are *equivalent* if they determine the same circular ordering around each vertex, and a *planar embedding* is an equivalence class of planar drawings. A graph is *embedded* when an embedding of it has been decided. A planar drawing partitions the plane into topologically connected regions, called *faces*. The unbounded face is the *outer face*, while the bounded faces are the *internal faces*. The outer face of a graph G is denoted by $f(G)$. A graph together with a planar embedding and a choice for its outer face is a *plane graph*. In a plane graph, *external* and *internal* vertices are defined as the vertices incident and not incident to the outer face, respectively. Sometimes the distinction is made between a *planar embedding* and *plane embedding*, where the former is an equivalence class of planar drawings and the latter is a planar embedding together with a choice for the outer face. The *dual graph* of an embedded planar graph G has a vertex for each face of G and has an edge (f,g) for each two faces f and g of G sharing an edge.

2.2 Maximality and Connectivity

A plane graph is *maximal* (or equivalently is a *triangulation*) when all its faces are delimited by 3-*cycles*, that is, by cycles of three vertices. A planar graph is *maximal* when it can be embedded as a triangulation. Algorithms for drawing planar graphs usually assume we are dealing with maximal planar graphs. In fact, any planar graph can be augmented to a maximal planar graph by adding some "dummy" edges to the graph. Then the algorithm can draw the maximal planar graph, and finally the inserted dummy edges can be removed, obtaining a drawing of the input graph.

A graph is *connected* if every pair of vertices is connected by a path. A graph with at least $k+1$ vertices is *k-connected* if removing any (at most) $k-1$ vertices leaves the graph connected; 3-connected, 2-connected, and 1-connected graphs are also called *triconnected*, *biconnected*, and *connected* graphs, respectively. A *separating cycle* is a cycle whose removal disconnects the graph.

2.3 Classes of Planar Graphs

A *tree* is a connected acyclic graph. A *leaf* in a tree is a node of degree 1. A *caterpillar* C is a tree such that the removal from C of all the leaves and of their incident edges turns C into a path, called the *backbone* of the caterpillar.

A *rooted tree* is a tree with one distinguished node called the *root*. In a rooted tree, each node v at distance (i.e., length of the shortest path) d from the root is the *child* of the only node at distance $d-1$ from the root to which v is connected. A *binary tree* (a *ternary tree*) is a rooted tree such that each node has at most two children (respectively, three children). Binary and ternary trees can be supposed to be rooted at any node of degree at most 2 and 3, respectively. The *height* of a rooted tree is the maximum number of nodes in any path from the root to a leaf. Removing a nonleaf node u from a tree disconnects the tree into connected components. Those containing children of u are the *subtrees* of u.

A *complete tree* is a rooted tree such that each nonleaf node has the same number of children and such that each leaf has the same distance from the root. Complete trees of degree 3 and 4 are also called *complete binary trees* and *complete ternary trees*, respectively.

A rooted tree is *ordered* of a clockwise order of the neighbors of each node (i.e., a planar embedding) is specified. In an ordered binary tree and in an ordered ternary tree, fixing a linear ordering of the children of the root defines the *left* and *right child* of a node, and the *left*, *middle*, and *right child* of a node, respectively. If the tree is ordered and binary (ternary), the subtrees rooted at the left and right child (at the left, middle, and right child) of a node u are the *left* and the *right subtree* of u (the *left*, the *middle*, and the *right subtree* of u), respectively. Removing a path P from a tree disconnects the tree into connected components. The ones containing children of nodes in P are the *subtrees* of P. If the tree is ordered and binary (ternary), then each component is a *left* or *right subtree* (a *left*, *middle*, or *right subtree*) of $\mathcal{P}$, depending on whether the root of such subtree is a left or right child (is a left, middle, or right child) of a node in $\mathcal{P}$, respectively.

An *outerplane graph* is a plane graph such that all the vertices are incident to the outer face. An *outerplanar embedding* is a planar embedding such that all the vertices are incident to the same face. An *outerplanar graph* is a graph that admits an outerplanar embedding. A *maximal outerplane graph* is an outerplane graph such that all its internal faces are delimited by cycles of three vertices. A *maximal outerplanar embedding* is an outerplanar embedding such that all its faces, except for the one to which all the vertices are incident, are delimited by cycles of three vertices. A *maximal outerplanar graph* is a graph that admits a maximal outerplanar embedding. Every outerplanar graph can be augmented to maximal by adding dummy edges to it.

If we do not consider the vertex corresponding to the outer face of G and its incident edges, then the *dual graph* of an outerplane graph G is a tree. Hence, when dealing with outerplanar graphs, we talk about the *dual tree* of an outerplanar graph (meaning the dual graph of an outerplane embedding of the outerplanar graph). The nodes of the dual tree of a maximal outerplane graph G have degree at most 3. Hence, the dual tree of G can be rooted to be a binary tree.

Series-parallel graphs are the graphs that can be inductively constructed as follows. An edge (u,v) is a series-parallel graph with *poles* u and v. Denote by u_i and v_i the poles of a series-parallel graph G_i. Then a *series composition* of a sequence $G_1, G_2, \dots, G_k$ of series-parallel graphs, with $k \geq 2$, constructs a series-parallel graph that has poles $u = u_1$ and $v = v_k$, that contains graphs G_i as subgraphs, and such that vertices v_i and u_{i+1} have been identified to be the same vertex, for each $i = 1, 2, \dots, k-1$. A *parallel composition* of a set $G_1, G_2, \dots, G_k$ of series-parallel graphs, with $k \geq 2$, constructs a series-parallel graph that has poles $u = u_1 = u_2 = \cdots = u_k$ and $v = v_1 = v_2 = \cdots = v_k$, that contains graphs G_i as subgraphs, and such that vertices $u_1, u_2, \dots, u_k$ (vertices $v_1, v_2, \dots, v_k$) have been identified to be the same vertex. A *maximal series-parallel graph* is such that all its series compositions construct a graph out of exactly two smaller series-parallel graphs G_1 and G_2, and such that all its parallel compositions have a component that is the edge between the two poles. Every series-parallel graph can be augmented to maximal by adding dummy edges to it. The *fan-out* of a series-parallel graph is the maximum number of components in a parallel composition.

A graph G is *bipartite* if its vertex set V can be partitioned into two subsets V_1 and V_2 so that every edge of G is incident to a vertex of V_1 and to a vertex of V_2. A *bipartite planar graph* is both bipartite and planar. A *maximal bipartite planar graph* admits a planar embedding in which all its faces have exactly four incident vertices. Every bipartite planar graph with at least four vertices can be augmented to maximal by adding dummy edges to it.

2.4 Drawing Standards

A *straight-line drawing* is a drawing such that each edge is represented by a segment. A *polyline drawing* is a drawing such that each edge is represented by a sequence of consecutive segments. The points in which two consecutive segments of the same edge touch are called *bends*. A *grid drawing* is a drawing such that vertices and bends have integer coordinates. An *orthogonal drawing* is a polyline drawing such that each edge is represented by a sequence of horizontal and vertical segments. A *convex drawing* (respectively, *strictly-convex drawing*) is a planar drawing such that each face is delimited by a convex polygon (respectively, strictly-convex polygon); that is, every interior angle of the drawing is at most 180° (respectively, less than 180°) and every exterior angle is at least 180° (respectively, more than 180°). An *order-preserving drawing* is a drawing such that the order of the edges incident to each vertex respects an order fixed in advance. An *upward drawing* (respectively, *strictly-upward drawing*) of a rooted tree is a drawing such that each edge is represented by a nondecreasing curve (respectively, increasing curve). A *visibility representation* is a drawing such that each vertex is represented by a horizontal segment $\sigma(u)$, each edge (u,v) is represented by a vertical segment connecting a point of $\sigma(u)$ with a point of $\sigma(v)$, and no two segments cross, except if they represent a vertex and one of its incident edges.

2.5 Area of a Drawing

The *bounding box* of a drawing is the smallest rectangle with sides parallel to the axes that contains the drawing completely. The *height* and *width* of a drawing are the height and width of its bounding box. The *area* of a drawing is the area of its bounding box. The *aspect ratio* of a drawing is the ratio between the maximum and the minimum of the height and width of the drawing. Observe that the concept of area of a drawing only makes sense once a *resolution rule* is fixed, i.e., a rule that does not allow vertices to be arbitrarily close (*vertex resolution rule*), or edges to be arbitrarily short (*edge resolution rule*). Without any such rules, one could just construct drawings with an arbitrarily small area. It is usually assumed in the literature that graph drawings in a small area have to be constructed on a grid. In fact, all the algorithms we will present in Sects. 3–6 and 8 assign integer coordinates to vertices. The assumption of constructing drawings on the grid is usually relaxed in the context of proximity drawings (hence in Sect. 7), where in fact it is assumed that no two vertices have distance less than one unit.

2.6 Directed Graphs and Planar Upward Drawings

A *directed acyclic graph* (*DAG* for short) is a graph whose edges are oriented and containing no cycle $(v_1,\ldots,v_n)$ such that edge (v_i,v_{i+1}) is directed from v_i to v_{i+1}, for $i=1,\ldots,n-1$, and edge (v_n,v_1) is directed from v_n to v_1. The *underlying graph* of a DAG G is the undirected graph obtained from G by removing the directions on its edges. An *upward drawing* of a DAG is such that each edge is represented by an increasing curve. An *upward planar drawing* is a drawing that is both upward and planar. An *upward planar DAG* is a DAG that admits an upward planar drawing. In a directed graph, the *outdegree* of a vertex is the number of edges leaving the vertex, and the *indegree* of a vertex is the number of edges entering the vertex. A *source* (respectively, *sink*) is a vertex with indegree zero (respectively, with outdegree zero). An *st-planar DAG* is a DAG with exactly one source s and one sink t that admits an upward planar embedding in which s and t are on the outer face. *Bipartite DAGs* and *directed trees* are DAGs whose underlying graphs are bipartite graphs and trees, respectively. A *series-parallel DAG* is a DAG that can be inductively constructed as follows. An edge (u,v) directed from u to v is a series-parallel DAG with *starting pole* u and *ending pole* v. Denote by u_i and v_i the starting and ending poles of a series-parallel DAG G_i, respectively. Then a *series composition* of a sequence $G_1,G_2,\ldots,G_k$ of series-parallel DAGs, with $k\geq 2$, constructs a series-parallel DAG that has starting pole $u=u_1$, that has ending pole $v=v_k$, that contains DAGs G_i as subgraphs, and such that vertices v_i and u_{i+1} have been identified to be the same vertex, for each $i=1,2,\ldots,k-1$. A *parallel composition* of a set $G_1,G_2,\ldots,G_k$ of series-parallel DAGs, with $k\geq 2$, constructs a series-parallel DAG that has starting pole $u=u_1=u_2=\cdots=u_k$, that has ending pole $v=v_1=v_2=\cdots=v_k$, that contains DAGs G_i as subgraphs, and such that vertices $u_1,u_2,\ldots,u_k$ (vertices $v_1,v_2,\ldots,v_k$) have been identified to be the same vertex. We remark that series-parallel DAGs are a subclass of the upward planar DAGs whose underlying graph is a sseries-parallel graph.

2.7 Proximity Drawings

A *Delaunay drawing* of a graph G is a straight-line drawing such that no three vertices are on the same line, no four vertices are on the same circle, and three vertices u, v, and z form a 3-cycle (u,v,z) in G if and only if the circle passing through u, v, and z in the drawing contains no vertex other than u, v, and z. A *Delaunay triangulation* is a graph that admits a Delaunay drawing.

The *Gabriel region* of two vertices x and y is the disk having segment $\overline{xy}$ as diameter. A *Gabriel drawing* of a graph G is a straight-line drawing of G having the property that two vertices x and y of the drawing are connected by an edge if and only if the Gabriel region of x and y does not contain any other vertex. A *Gabriel graph* is a graph admitting a Gabriel drawing.

A *relative neighborhood drawing* of a graph G is a straight-line drawing such that two vertices x and y are adjacent if and only if there is no vertex whose distance to both x and y is less than the distance between x and y. A *relative neighborhood graph* is a graph admitting a relative neighborhood drawing.

A *nearest-neighbor drawing* of a graph G is a straight-line drawing of G such that each vertex has a unique closest vertex and such that two vertices x and y of the drawing are connected by an edge if and only if x is the vertex of G closest to y, or vice versa. A *nearest-neighbor graph* is a graph admitting a nearest-neighbor drawing.

A β*-drawing* is a straight-line drawing of G having the property that two vertices x and y of the drawing are connected by an edge if and only if the β-region of x and y does not contain any other vertex. The β*-region* of x and y is the line segment $\overline{xy}$ if $\beta = 0$, it is the intersection of the two closed disks of radius $d(x,y)/(2\beta)$ passing through both x and y if $0 < \beta < 1$, it is the intersection of the two closed disks of radius $d(x,y)/(2\beta)$ that are centered on the line through x and y and that respectively pass through x and through y if $1 \leq \beta < \infty$, and it is the closed infinite strip perpendicular to the line segment $\overline{xy}$ if $\beta = \infty$.

Weak proximity drawings are such that there is no geometric requirement on the pairs of vertices not connected by an edge. For example, a *weak Gabriel drawing* of a graph G is a straight-line drawing of G having the property that if two vertices x and y of the drawing are connected by an edge, then the Gabriel region of x and y does not contain any other vertex, while there might exist two vertices whose Gabriel region is empty and that are not connected by an edge.

A *Euclidean minimum spanning tree* T of a set P of points is a tree spanning the points in P (that is, the nodes of T coincide with the points of P and no "Steiner points" are allowed) and having minimum total edge length.

A *greedy drawing* of a graph G is a straight-line drawing of G such that, for every pair of nodes u and v, there exists a *distance-decreasing path*, where a path $(v_0, v_1, \ldots, v_m)$ is distance-decreasing if $d(v_i, v_m) < d(v_{i-1}, v_m)$, for $i = 1, \ldots, m$, where $d(p,q)$ denotes the Euclidean distance between two points p and q.

For more about proximity drawings, see Chap. 7 in [115].

2.8 Clustered Graphs and c-Planar Drawings

A *clustered graph* is a pair $C(G,T)$, where G is a graph, called an *underlying graph*, and T is a rooted tree, called an *inclusion tree*, such that the leaves of T are the vertices of G. Each internal node ν of T corresponds to the subset of vertices of G, called a *cluster*, that are the leaves of the subtree of T rooted at ν. A clustered graph $C(G,T)$ is *c-connected* if each cluster induces a connected subgraph of G; it is *non-c-connected* otherwise.

A *drawing* Γ of a clustered graph $C(G,T)$ consists of a drawing of G (each vertex is a point in the plane and each edge is a Jordan curve between its end vertices) and of a representation of each node μ of T as a simple closed region containing all and only the vertices that belong to μ. A drawing is *c-planar* if it has no edge crossings (i.e., the drawing of the underlying graph is planar), no edge–region crossings (i.e., an edge intersects the boundary of a cluster at most once), and no region–region crossings (i.e., no two cluster boundaries cross).

A *c-planar embedding* is an equivalence class of *c*-planar drawings of C, where two *c*-planar drawings are equivalent if they have the same order of the edges incident to each vertex and the same order of the edges incident to each cluster.

3 Straight-Line Drawings

In this section, we discuss algorithms and bounds for constructing small-area planar straight-line drawings of planar graphs and their subclasses. In Sect. 3.1 we deal with general planar graphs, in Sect. 3.2 we deal with 4-connected and bipartite graphs, in Sect. 3.3 we deal with series-parallel graphs, in Sect. 3.4 we deal with outerplanar graphs, and in Sect. 3.5 we deal with trees. Table 1 summarizes the best-known area bounds for straight-line planar drawings of planar graphs and their subclasses. Observe that the lower bounds of the table that refer to general planar graphs, 4-connected planar graphs, and bipartite planar graphs have been obtained considering plane graphs.

Table 1 A table summarizing the area requirements for straight-line planar drawings of several classes of planar graphs

	Upper bound	Refs.	Lower bound	Refs.
General planar graphs	$\frac{8n^2}{9}+O(n)$	[20, 37, 110]	$\frac{4n^2}{9}-O(n)$	[37, 69, 100, 121]
4-Connected planar graphs	$\lfloor\frac{n}{2}\rfloor\times(\lceil\frac{n}{2}\rceil-1)$	[99]	$\lfloor\frac{n}{2}\rfloor\times(\lceil\frac{n}{2}\rceil-1)$	[99]
Bipartite planar graphs	$\lfloor\frac{n}{2}\rfloor\times(\lceil\frac{n}{2}\rceil-1)$	[14]	$\lfloor\frac{n}{2}\rfloor\times(\lceil\frac{n}{2}\rceil-1)$	[14]
Series-parallel graphs	$O(n^2)$	[37, 110, 126]	$\Omega(n2^{\sqrt{\log n}})$	[67]
Outerplanar graphs	$O(n^{1.48})$	[42]	$\Omega(n)$	Trivial
Trees	$O(n\log n)$	[34]	$\Omega(n)$	Trivial

Notice that 4-connected planar graphs have been studied only with the additional constraint of having at least four vertices on the outer face.

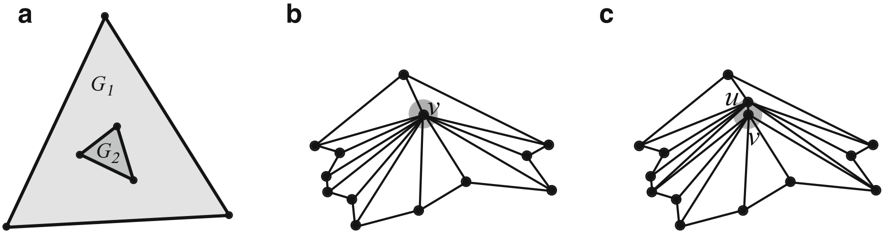


Fig. 1 (**a**) Induction in Fáry's algorithm if G contains a separating 3-cycle. (**b**, **c**) Induction in Fáry's algorithm if G contains no separating 3-cycle

3.1 General Planar Graphs

In this section, we discuss algorithms and bounds for constructing small-area planar straight-line drawings of general planar graphs. Observe that in order to derive bounds on the area requirements of general planar graphs, it suffices to restrict the attention to maximal planar graphs, as every planar graph can be augmented to maximal by the insertion of "dummy" edges. Moreover, such an augmentation can be performed in linear time [108].

We start by proving that every plane graph admits a planar straight-line drawing [113, 122]. The simplest and most elegant proof of such a statement is, in our opinion, the one presented by Fáry in 1948 [59].

Fáry's algorithm works by induction on the number n of vertices of the plane graph G; namely, the algorithm inductively assumes that a straight-line planar drawing of G can be constructed with the further constraint that the outer face $f(G)$ is drawn as an arbitrary triangle Δ. The inductive hypothesis is trivially satisfied when $n = 3$. If $n > 3$, then two cases are possible. In the first case, G contains a separating 3-cycle c. Then let G_1 (respectively, G_2) be the graph obtained from G by removing all the vertices internal to c (respectively, external to c). Both G_1 and G_2 have less than n vertices; hence, the inductive hypothesis applies first to construct a straight-line planar drawing Γ_1 of G_1 in which $f(G_1)$ is drawn as an arbitrary triangle Δ, and second to construct a straight-line planar drawing Γ_2 of G_2 in which $f(G_2)$ is drawn as $\Delta(c)$, where $\Delta(c)$ is the triangle representing c in Γ_1 (see Fig. 1a). Thus, a straight-line drawing Γ of G in which $f(G)$ is represented by Δ is obtained. In the second case, G does not contain any separating 3-cycle; i.e., G is 4-connected. Then, consider any internal vertex u of G and consider any neighbor v of u. Construct an $(n-1)-$vertex plane graph G' by removing u and all its incident edges from G, and by inserting "dummy" edges between v and all the neighbors of u in G, except for the two vertices v_1 and v_2 forming faces with u and v. The graph G' is simple, as G contains no separating 3-cycle. Hence, the inductive hypothesis applies to construct a straight-line planar drawing Γ' of G' in which $f(G')$ is drawn as Δ. Further, dummy edges can be removed and vertex u can be introduced in Γ' together with its incident edges, without altering the planarity of Γ'. In fact, u can be

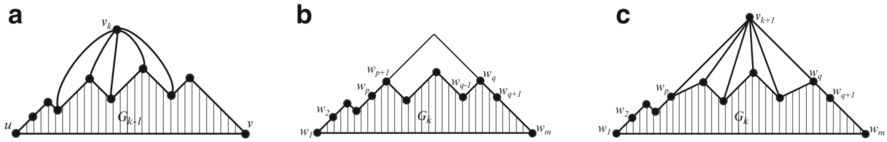


Fig. 2 (**a**) The canonical ordering of a maximal plane graph G. (**b**) The drawing of G_k constructed by the algorithm of de Fraysseix et al. (**c**) The drawing of G_{k+1} constructed by the algorithm of de Fraysseix et al.

placed at a suitable point in the interior of a small disk centered at v, thus obtaining a straight-line drawing Γ of G in which $f(G)$ is represented by Δ (see Fig. 1b, c).

The first algorithms for constructing planar straight-line grid drawings of planar graphs in a polynomial area were presented (50 years later than Fáry's algorithm!) by de Fraysseix et al. [36, 37] and, simultaneously and independently, by Schnyder [110]. The approaches of the two algorithms, which we sketch below, are today still the basis of every algorithm to construct planar straight-line grid drawings of triangulations.

The algorithm by de Fraysseix et al. [36, 37] relies on two main ideas.

First, any n-vertex maximal plane graph G admits a total ordering σ of its vertices, called a *canonical ordering*, such that (see Fig. 2a): (a) the subgraph G_k of G induced by the first k vertices in σ is biconnected, for each $k = 3, \dots, n$; and (b) the kth vertex in σ lies in the outer face of G_{k-1}, for each $k = 4, \dots, n$.

Second, a straight-line drawing of an n-vertex maximal plane graph G can be constructed starting from a drawing of the 3-cycle induced by the first three vertices in a canonical ordering σ of G and incrementally adding vertices to the partially constructed drawing in the order defined by σ. To construct the drawing of G one vertex at a time, the algorithm maintains the invariant that the outer face of G_k is drawn as a sequence of segments having slopes equal to either $45°$ or $-45°$. When the next vertex v_{k+1} in σ is added to the drawing of G_k to construct a drawing of G_{k+1}, a subset of the vertices of G_k undergoes a horizontal shift that allows for v_{k+1} to be introduced in the drawing while still maintaining the invariant that the outer face of G_{k+1} is drawn as a sequence of segments having slopes equal to either $45°$ or $-45°$ (see Fig. 2b, c).

The area of the constructed drawings is $(2n-4) \times (n-2)$. The described algorithm has been proposed by de Fraysseix et al. together with an $O(n \log n)$-time implementation. The authors conjectured that its complexity could be improved to $O(n)$. This bound was in fact achieved a few years later by Chrobak and Payne in [29].

The ideas behind the algorithm by Schnyder [110] are totally different from the ones of de Fraysseix et al. In fact, Schnyder's algorithm constructs the drawing by determining the coordinates of all the vertices in one shot. The algorithm relies on results concerning planar graph embeddings that are indeed less intuitive than the canonical ordering of a plane graph used by de Fraysseix et al.

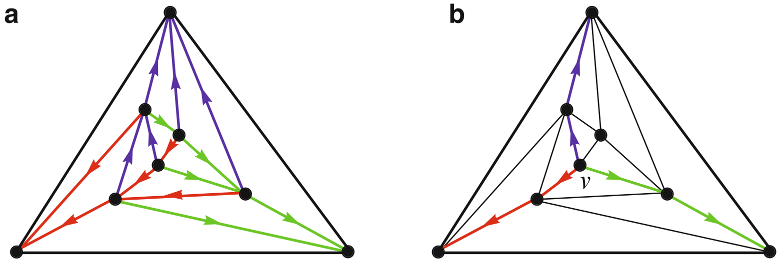


Fig. 3 (**a**) A realizer for a plane graph G. (**b**) Paths $P_1(v)$, $P_2(v)$, and $P_3(v)$ (represented by *green*, *red*, and *blue edges*, respectively) and regions $R_1(v)$, $R_2(v)$, and $R_3(v)$ [delimited by $P_1(v)$, $P_2(v)$, and $P_3(v)$, and by the edges incident to the outer face of G]

First, Schnyder introduces the concept of the *barycentric representation* of a graph G as an injective function $v \in V(G) \rightarrow (x(v), y(v), z(v))$ such that $x(v) + y(v) + z(v) = 1$, for all vertices $v \in V(G)$, and such that, for each edge $(u,v) \in E(G)$ and each vertex $w \notin \{u,v\}$, $x(u) < x(w)$ and $x(v) < x(w)$ hold, or $y(u) < y(w)$ and $y(v) < y(w)$ hold, or $z(u) < z(w)$ and $z(v) < z(w)$ hold. Schnyder proves that, given any graph G, given any barycentric representation $v \rightarrow (x(v), y(v), z(v))$ of G, and given any three noncollinear points α, β, and γ in the three-dimensional space, the mapping $f : v \in V(G) \rightarrow v_1\alpha + v_2\beta + v_3\gamma$ is a straight-line planar embedding of G in the plane spanned by α, β, and γ.

Second, Schnyder introduces the concept of a *realizer* of G as an orientation and a partition of the interior edges of a plane graph G into three sets T_1, T_2, and T_3 such that (a) the set of edges in T_i, for each $i = 1,2,3$, is a tree spanning all the internal vertices of G and exactly one external vertex; (b) all the edges of T_i are directed toward this external vertex, which is the root of T_i; (c) the external vertices belonging to T_1, to T_2, and to T_3 are distinct and appear in counterclockwise order on the border of the outer face of G; and (d) the counterclockwise order of the edges incident to v is as follows: Leaving T_1, entering T_3, leaving T_2, entering T_1, leaving T_3, and entering T_2. Figure 3a illustrates a realizer for a plane graph G. Trees T_1, T_2, and T_3 are sometimes called *Schnyder woods*.

Third, Schnyder describes how to get a barycentric representation of a plane graph G starting from a realizer of G; this is essentially done by looking, for each vertex $v \in V(G)$, at the paths $P_i(v)$—they are the only paths composed entirely of edges of T_i connecting v to the root of T_i (see Fig. 3b)—and by counting the number of faces or the number of vertices in the regions $R_1(v)$, $R_2(v)$, and $R_3(v)$ that are defined by $P_1(v)$, $P_2(v)$, and $P_3(v)$. The area of the constructed drawings is $(n-2) \times (n-2)$.

Schnyder's upper bound has been unbeaten for almost 20 years. Only recently, Brandenburg [20] proposed an algorithm for constructing planar straight-line drawings of triangulations in $\frac{8n^2}{9} + O(n)$ area. Such an algorithm is based on a geometric refinement of the de Fraysseix et al. [36, 37] algorithm combined with some topological properties of planar triangulations due to Bonichon et al. [18], which will be discussed in Sect. 4.

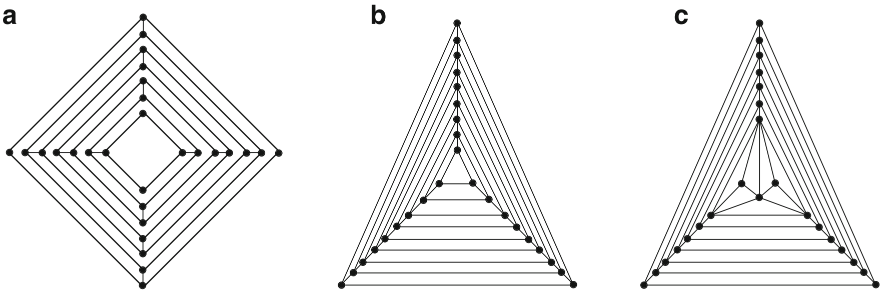


Fig. 4 (**a**) A graph [121] requiring quadratic area in any straight-line and polyline drawing. (**b**) A graph [37] requiring $\left(\frac{2n}{3}-1\right)\times\left(\frac{2n}{3}-1\right)$ area in any straight-line and polyline drawing. (**c**) A graph [69] requiring $\frac{4n^2}{9}-\frac{2n}{3}$ area in any straight-line drawing

A quadratic-area upper bound for straight-line planar drawings of plane graphs is asymptotically optimal. In fact, almost 10 years before the publication of such algorithms, Valiant observed in [121] that there exist n-vertex plane graphs (see Fig. 4a) requiring $\Omega(n^2)$ area in any straight-line planar drawing (in fact, in every polyline planar drawing). It was then proved by de Fraysseix et al. in [37] that *nested triangles graphs* (see Fig. 4b) require $\left(\frac{2n}{3}-1\right)\times\left(\frac{2n}{3}-1\right)$ area in any straight-line planar drawing (in fact, in every polyline planar drawing). Such a lower bound was only recently improved to $\frac{4n^2}{9}-\frac{2n}{3}$ by Frati and Patrignani [69], for all n multiples of 3 (see Fig. 4c), and then by Mondal et al. [100] to $\left\lfloor\frac{2n}{3}-1\right\rfloor\times\left\lfloor\frac{2n}{3}\right\rfloor$, for all $n\geq 6$.

However, the following remains open:

Open Problem 1. *Close the gap between the* $\frac{8n^2}{9}+O(n)$ *upper bound and the* $\frac{4n^2}{9}-O(n)$ *lower bound for the area requirements of straight-line drawings of plane graphs.*

3.2 4-Connected and Bipartite Planar Graphs

In this section, we discuss algorithms and bounds for constructing planar straight-line drawings of 4-connected and bipartite planar graphs. Such different families of graphs are discussed in the same section since the best-known upper bound for the area of bipartite planar graphs was obtained using an augmentation to 4-connectivity.

Concerning 4-connected plane graphs, tight bounds are known for the area requirements of planar straight-line drawings if the graph has at least four vertices incident to the outer face. Namely, Miura et al. proved in [99] that every such a graph has a planar straight-line drawing in $(\lceil\frac{n}{2}\rceil-1)\times(\lfloor\frac{n}{2}\rfloor)$ area, improving upon

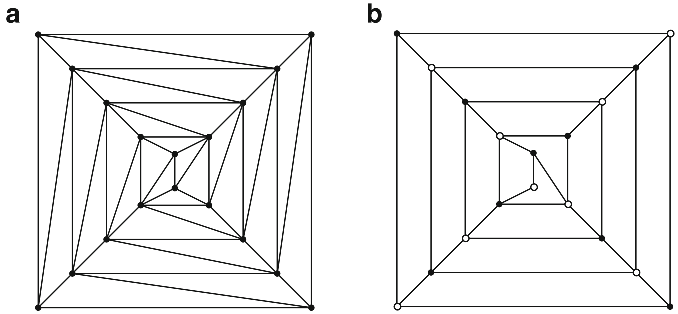
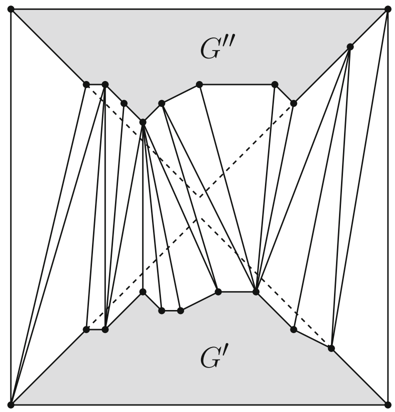


Fig. 5 (**a**) A 4-connected plane graph requiring $(\lceil \frac{n}{2} \rceil - 1) \times (\lfloor \frac{n}{2} \rfloor)$ area in any straight-line planar drawing. (**b**) A bipartite plane graph requiring $(\lceil \frac{n}{2} \rceil - 1) \times (\lfloor \frac{n}{2} \rfloor)$ area in any straight-line planar drawing

Fig. 6 The algorithm by Miura et al. to construct straight-line drawings of 4-connected plane graphs [99]



previous results of He [82]. The authors show that this bound is tight, by exhibiting a class of 4-connected plane graphs with four vertices incident to the outer face requiring $(\lceil \frac{n}{2} \rceil - 1) \times (\lfloor \frac{n}{2} \rfloor)$ area (see Fig. 5a).

The algorithm of Miura et al. divides the input 4-connected plane graph G into two graphs G' and G'' with the same number of vertices. This is done by performing a *4-canonical ordering* [89] of G. The graph G' (G'', respectively) is then drawn inside an isosceles right triangle Δ' (respectively, Δ'') whose width is $\frac{n}{2} - 1$ and whose height is half of its width. To construct such drawings of G' and G'', Miura et al. designed an algorithm that is similar to the algorithm by de Fraysseix et al. [37]. In the drawings produced by their algorithm, the slopes of the edges incident to the outer faces of G' and G'' have absolute value at most 45°. The drawing of G'' is then rotated by 180° and placed on top of the drawing of G'. This allows for drawing the edges connecting G' with G'' without creating crossings. Figure 6 depicts the construction of Miura et al.'s algorithm.

As far as we know, no bound better than the one for general plane graphs is known for 4-connected plane graphs (possibly having three vertices incident to the outer face); hence, the following is open:

Open Problem 2. *Close the gap between the* $\frac{8n^2}{9} + O(n)$ *upper bound and the* $\frac{n^2}{4} - O(n)$ *lower bound for the area requirements of straight-line drawings of* 4-*connected plane graphs.*

Biedl and Brandenburg [14] show how to construct planar straight-line drawings of bipartite planar graphs in $(\lceil \frac{n}{2} \rceil - 1) \times (\lfloor \frac{n}{2} \rfloor)$ area. To achieve such a bound, they exploit a result of Biedl et al. [16] stating that all planar graphs without separating triangles—except those "containing a star" (see [14] and observe that in this case a star is not just a vertex plus some incident edges)—can be augmented to 4-connected by the insertion of dummy edges; once such an augmentation is done, Biedl and Brandenburg use the algorithm of Miura et al. [99] to draw the resulting 4-connected plane graph. In order to be able to use Miura et al.'s algorithm, Biedl and Brandenburg prove that no bipartite plane graph "contains a star" and that Miura et al.'s algorithm works more in general for plane graphs that become 4-connected if an edge is added to them. The upper bound of Biedl and Brandenburg is tight, as the authors show a bipartite plane graph requiring $(\lceil \frac{n}{2} \rceil - 1) \times (\lfloor \frac{n}{2} \rfloor)$ area in any straight-line planar drawing (see Fig. 5b).

3.3 Series-Parallel Graphs

In this section, we discuss algorithms and bounds for constructing small-area planar straight-line drawings of series-parallel graphs.

No subquadratic-area upper bound is known for constructing small-area planar straight-line drawings of series-parallel graphs. The best-known quadratic upper bound for straight-line drawings is provided in [126].

In [67] Frati proved that series-parallel graphs exist that require $\Omega(n2^{\sqrt{\log n}})$ area in any straight-line or polyline grid drawing. Such a result is achieved in two steps. In the first one, an $\Omega(n)$ lower bound for the maximum between the height and width of any straight-line or polyline grid drawing of $K_{2,n}$ is proved, thus answering a question of Felsner et al. [60] and improving upon previous results of Biedl et al. [15]. In the second one, an $\Omega(2^{\sqrt{\log n}})$ lower bound for the minimum between the height and width of any straight-line or polyline grid drawing of certain series-parallel graphs is proved.

The proof that $K_{2,n}$ requires $\Omega(n)$ height or width in any straight-line or polyline drawing has several ingredients. First, a simple "optimal" drawing algorithm for $K_{2,n}$ is exhibited; that is, an algorithm is presented that computes a drawing of $K_{2,n}$ inside an arbitrary convex polygon if such a drawing exists. Second, the drawings constructed by the mentioned algorithm inside a rectangle are studied. Such a study reveals that the slopes of the segments representing the edges of $K_{2,n}$ have a strong

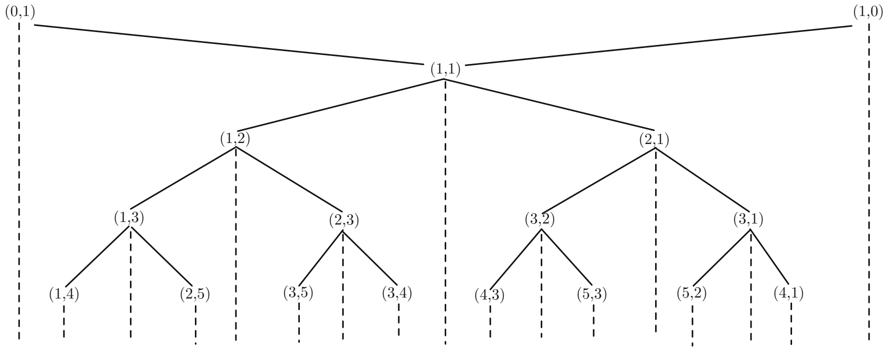


Fig. 7 The Stern–Brocot tree is a tree containing all the pairs of relatively prime numbers

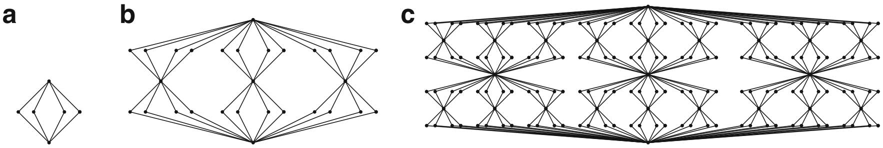


Fig. 8 The inductive construction of series-parallel graphs requiring $\Omega(2^{\sqrt{\log n}})$ height and width in any straight-line or polyline grid drawing

relationship with the relatively prime numbers as ordered in the *Stern-Brocot* tree (see [21, 114] and Fig. 7). Such a relationship leads to deriving some arithmetical properties of the lines passing through infinite grid points in the plane and to achieve the $\Omega(n)$ lower bound.

The results on the area requirements of $K_{2,n}$ are then used to construct series-parallel graphs (shown in Fig. 8) out of several copies of $K_{2,2^{\sqrt{\log n}}}$ and to prove that such a graph requires $\Omega(2^{\sqrt{\log n}})$ height and width in any straight-line or polyline grid drawing.

As no subquadratic-area upper bound is known for straight-line planar drawings of series-parallel graphs, the following is open.

Open Problem 3. *Close the gap between the $O(n^2)$ upper bound and the $\Omega(n2^{\sqrt{\log n}})$ lower bound for the area requirements of straight-line drawings of series-parallel graphs.*

Related to the above problem, Wood (Private Communication, 2008) conjectures the following: Let $p_1, \dots, p_k$ be positive integers. Let $G(p_1)$ be the graph obtained from K_3 by adding p_1 new vertices adjacent to v and w for each edge (v,w) of K_3. For $k \geq 2$, let $G(p_1, p_2, \dots, p_k)$ be the graph obtained from $G(p_1, p_2, \dots, p_{k-1})$ by adding p_k new vertices adjacent to v and w for each edge (v,w) of $G(p_1, p_2, \dots, p_{k-1})$. Observe that $G(p_1, p_2, \dots, p_k)$ is a sseries-parallel graph.

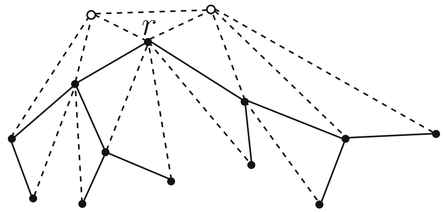


Fig. 9 A star-shaped drawing Γ of a binary tree T (with *thick edges* and *black vertices*). The *dashed edges* and *white vertices* augment Γ into a straight-line drawing of the outerplanar graph to which T is dual

Conjecture 1 (D. R. Wood). Every straight-line grid drawing of $G(p_1, p_2, \ldots, p_k)$ requires $\Omega(n^2)$ area for some choice of k and $p_1, p_2, \ldots, p_k$.

3.4 Outerplanar Graphs

In this section, we discuss algorithms and bounds for constructing small-area planar straight-line drawings of outerplanar graphs.

The first nontrivial bound appeared in [74], where Garg and Rusu proved that every outerplanar graph with maximum degree d has a straight-line drawing with $O(dn^{1.48})$ area. Such a result is achieved by means of an algorithm that works by induction on the dual tree T of the outerplanar graph G. Namely, the algorithm finds a path P in T, it removes from G the subgraph G_P that has P as a dual tree, it inductively draws the outerplanar graphs that are disconnected by such a removal, and it puts all the drawings of such outerplanar graphs together with a drawing of G_P, obtaining a drawing of the whole outerplanar graph.

The first subquadratic-area upper bound for straight-line drawings of outerplanar graphs has been proved by Di Battista and Frati in [42]. The result in [42] uses the following ingredients. First, it is shown that the dual binary tree T of a maximal outerplanar graph G is a subgraph of G itself. Second, a restricted class of straight-line drawings of binary trees, called *star-shaped drawings*, is defined. Star-shaped drawings are straight-line drawings in which special visibility properties among the nodes of the tree are satisfied (see Fig. 9). Namely, if a tree T admits a star-shaped drawing Γ, then the edges that augment T into G can be drawn in Γ without creating crossings, thus resulting in a straight-line planar drawing of G. Third, an algorithm is shown to construct a star-shaped drawing of any binary tree T in $O(n^{1.48})$ area. Such an algorithm works by induction on the number of nodes of T. Figure 10 depicts two inductive cases of such a construction, making use of a strong combinatorial decomposition of ordered binary trees introduced by Chan in [24] (discussed in Sect. 3.5).

Frati used in [64] the same approach of [42], together with a different geometric construction (shown in Fig. 11), to prove that every outerplanar graph with degree d has a straight-line drawing with $O(dn \log n)$ area.

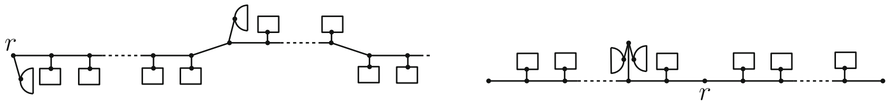


Fig. 10 Two inductive cases of the algorithm to construct star-shaped drawings of binary trees yielding an $O(n^{1.48})$ upper bound for straight-line drawings of outerplanar graphs. The *rectangles* and the *half-circles* represent subtrees recursively drawn by the construction on the *right* and on *left part* of the figure, respectively

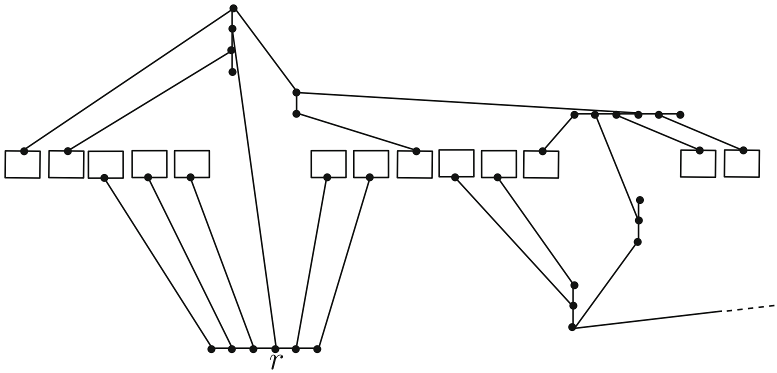


Fig. 11 The inductive construction of star-shaped drawings of binary trees yielding an $O(dn\log n)$ upper bound for straight-line drawings of outerplanar graphs with degree d. The *rectangles* represent recursively constructed star-shaped drawings of subtrees

As far as we know, no superlinear-area lower bound is known for straight-line drawings of outerplanar graphs. In [12] Biedl defined a class of outerplanar graphs, called *snowflake graphs*, and conjectured that such graphs require $\Omega(n\log n)$ area in any straight-line or polyline drawing. However, Frati disproved such a conjecture in [64] by exhibiting $O(n)$ area straight-line drawings of snowflake graphs. In the same paper, he conjectured that an $O(n\log n)$ area upper bound for straight-line drawings of outerplanar graphs cannot be achieved by squeezing the drawing along one coordinate direction, as stated in the following.

Conjecture 2 (F. Frati). There exist n-vertex outerplanar graphs for which, for any straight-line drawing in which the longest side of the bounding box is $O(n)$, the smallest side of the bounding box is $\omega(\log n)$.

The following problem remains wide open.

Open Problem 4. *Close the gap between the $O(n^{1.48})$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line drawings of outerplanar graphs.*

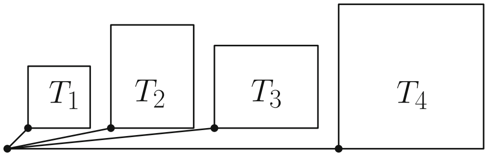


Fig. 12 Inductive construction of a straight-line drawing of a tree in $O(n \log n)$ area

3.5 Trees

In this section, we present algorithms and bounds for constructing planar straight-line drawings of trees.

The best bound for constructing general trees is, as far as we know, the $O(n \log n)$ area upper bound provided by a simple modification of the *hv*-drawing algorithm of Crescenzi et al. [34]. Such an algorithm proves that a straight-line drawing of any tree T in $O(n) \times O(\log n)$ area can be constructed with the further constraint that the root of T is placed at the bottom-left corner of the bounding box of the drawing. If T has one node, such a drawing is trivially constructed. If T has more than one node, then let $T_1, \dots, T_k$ be the subtrees of T, where we assume, w.l.o.g., that T_k is the subtree of T with the greatest number of nodes. Then the root of T is placed at $(0,0)$, the subtrees $T_1, \dots, T_{k-1}$ are placed one beside the other, with the bottom side of their bounding boxes on the line $y = 1$, and T_k is placed beside the other subtrees, with the bottom side of its bounding box on the line $y = 0$. The width of the drawing is clearly $O(n)$, while its height is $h(n) = \max\{h(n-1), 1 + h(n/2)\} = O(\log n)$, where $h(n)$ denotes the maximum height of a drawing of an n-node tree constructed by the algorithm. See Fig. 12 for an illustration of such an algorithm. Interestingly, no superlinear-area lower bound is known for the area requirements of straight-line drawings of trees.

For the special case of bounded-degree trees, linear-area bounds have been achieved. In fact, Garg and Rusu presented an algorithm to construct straight-line drawings of binary trees in $O(n)$ area [73] and an algorithm to construct straight-line drawings of trees with degree $O(\sqrt{n})$ in $O(n)$ area [72]. Both algorithms rely on the existence of simple *separators* for bounded degree trees. Namely, every binary tree T has a *separator edge*, that is, an edge whose removal disconnects T into two trees both having at most $2n/3$ vertices [121], and every degree-d tree T has a vertex whose removal disconnects T into at most d trees, each having at most $n/2$ nodes [72]. Such separators are exploited by Garg and Rusu to design quite complex inductive algorithms that achieve linear-area bounds and an optimal aspect ratio.

The following problem remains open.

Open Problem 5. *Close the gap between the $O(n \log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line drawings of trees.*

A lot of attention has been devoted to studying the area requirements of straight-line drawings of trees satisfying additional constraints. Table 2 summarizes the best-known area bounds for various kinds of straight-line drawings of trees.

Table 2 Summary of the best-known area bounds for straight-line drawings of trees

	Ord. Pres.	Upw.	Str. Upw.	Orth.	Upper bound	Refs.	Lower bound	Refs.
Binary					$O(n)$	[73]	$\Omega(n)$	Trivial
Binary	✓				$O(n\log\log n)$	[71]	$\Omega(n)$	Trivial
Binary		✓			$O(n\log\log n)$	[112]	$\Omega(n)$	Trivial
Binary			✓		$O(n\log n)$	[34]	$\Omega(n\log n)$	[34]
Binary	✓		✓		$O(n\log n)$	[71]	$\Omega(n\log n)$	[34]
Binary				✓	$O(n\log\log n)$	[23, 112]	$\Omega(n)$	Trivial
Binary		✓		✓	$O(n\log n)$	[23, 34]	$\Omega(n\log n)$	[23]
Binary	✓			✓	$O(n^{1.5})$	[66]	$\Omega(n)$	Trivial
Ternary				✓	$O(n^{1.631})$	[66]	$\Omega(n)$	Trivial
Ternary	✓			✓	$O(n^2)$	[66]	$\Omega(n^2)$	[66]
General					$O(n\log n)$	[34]	$\Omega(n)$	Trivial
General	✓				$O(n\log n)$	[71]	$\Omega(n)$	Trivial
General		✓			$O(n\log n)$	[34]	$\Omega(n)$	Trivial
General			✓		$O(n\log n)$	[34]	$\Omega(n\log n)$	[34]
General	✓		✓		$O(n4^{\sqrt{2\log n}})$	[24]	$\Omega(n\log n)$	[34]

"Ord. Pres.," "Upw.," "Str. Upw.," and "Orth." stand for order-preserving, upward, strictly upward, and orthogonal, respectively.

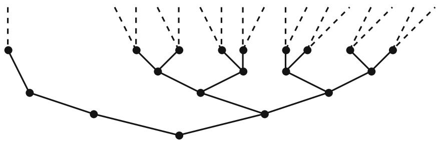
Concerning *straight-line upward drawings*, the illustrated algorithm of Crescenzi et al. [34] achieves the best-known upper bound of $O(n\log n)$. For trees with constant degree, Shin et al. prove in [112] that upward straight-line drawings in $O(n\log\log n)$ area can be constructed. Their algorithm is based on nice inductive geometric constructions and suitable tree decompositions. No superlinear-area lower bound is known, neither for binary nor for general trees; hence, the following are open.

Open Problem 6. *Close the gap between the $O(n\log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of upward straight-line drawings of trees.*

Open Problem 7. *Close the gap between the $O(n\log\log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of upward straight-line drawings of binary trees.*

Concerning *straight-line strictly upward drawings*, tight bounds are known. In fact, the algorithm of Crescenzi et al. [34] can be suitably modified in order to obtain strictly upward drawings (instead of aligning the subtrees of the root with their bottom sides on the same horizontal line, it is sufficient to align them with their left sides on the same vertical line). The same authors also showed a binary tree T^* requiring $\Omega(n\log n)$ area in any strictly upward drawing, and hence their bound is tight. The tree T^*, which is shown in Fig. 13, is composed of a path with $\Omega(n)$ nodes [forcing the height of the drawing to be $\Omega(n)$] and of a complete binary tree with $\Omega(n)$ nodes [forcing the width of the tree to be $\Omega(\log n)$].

Fig. 13 A binary tree T^* requiring $\Omega(n \log n)$ area in any strictly upward drawing



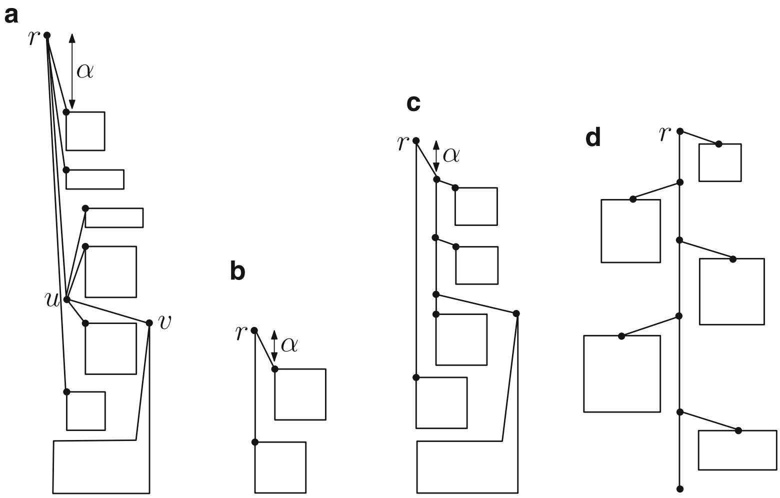


Fig. 14 (**a**) The inductive construction of a straight-line order-preserving drawing of a tree in $O(n \log n)$ area. (**b**, **c**) The inductive construction of a straight-line strictly upward order-preserving drawing of a binary tree in $O(n \log n)$ area. The construction in (**b**) [respectively, in (**c**)] refers to the case in which the *left* (respectively, the *right*) subtree of r contains more nodes than the *right* (respectively, the *left*) subtree of r. (**d**) The geometric construction of the algorithm of Chan

Concerning *straight-line order-preserving drawings*, Garg and Rusu have shown in [71] how to obtain an $O(n \log n)$ area upper bound for general trees. The algorithm of Garg and Rusu inductively assumes that an α-*drawing* of a tree T can be constructed; that is, a straight-line order-preserving drawing of T can be constructed with the further constraints that the root r of T is on the upper left corner of the bounding box of the drawing, that the children of r are placed on the vertical line one unit to the right of r, and that the vertical distance between r and any other node of T is at least α. Refer to Fig. 14a. To construct a drawing of T, the algorithm considers inductively constructed drawings of all the subtrees rooted at the children of r, except for the node u that is the root of the subtree of r with the greatest number of nodes, and place such drawings one unit to the right of r, with their left side aligned. Further, the algorithm considers inductively constructed drawings of all the subtrees rooted at the children of u, except for the node v that is the root of the subtree of u with the greatest number of nodes, and place such drawings two units to the right of r, with their left side aligned. Finally, the subtree rooted at v is inductively drawn, and then the drawing is reflected and placed with its left side on the same vertical line as r. Thus, the height of the drawing is clearly $O(n)$, while its width is $w(n) = \max\{w(n-1), 3 + w(n/2)\} = O(\log n)$, where $w(n)$ denotes the maximum width of a drawing of an n-node tree constructed by the algorithm. Garg and Rusu also show how to combine their described result with a decomposition scheme of binary trees due to Chan et al. [23] to obtain $O(n \log \log n)$ area straight-line order-preserving drawings of binary trees. As no superlinear lower bound is known for the area requirements of straight-line order-preserving drawings of trees, the following problems remain open.

Open Problem 8. *Close the gap between the $O(n \log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line order-preserving drawings of trees.*

Open Problem 9. *Close the gap between the $O(n \log \log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line order-preserving drawings of binary trees.*

Concerning *straight-line strictly upward order-preserving drawings*, Garg and Rusu have shown in [71] how to obtain an $O(n \log n)$ area upper bound for binary trees. Observe that such an upper bound is still matched by the described $\Omega(n \log n)$ lower bound of Crescenzi et al. [34]. The algorithm of Garg and Rusu, shown in Figs. 14b, c, is similar to their described algorithm for constructing straight-line order-preserving drawings of trees. The results of Garg and Rusu improved upon previous results by Chan in [24]. In [24], the author proved that every binary tree admits a straight-line strictly upward order-preserving drawing in $O(n^{1+\varepsilon})$ area, for any constant $\varepsilon > 0$. In the same paper, the author proved the best-known upper bound for the area requirements of straight-line strictly upward order-preserving drawings of trees, namely, $O(n 4^{\sqrt{2 \log n}})$. The approach of Chan consists of using very simple geometric constructions together with nontrivial tree decompositions. The simplest geometric construction discussed by Chan consists of selecting a path P in the input tree T, of drawing P on a vertical line l, and of inductively constructing drawings of the subtrees of P to be placed to the left and right of l (see Fig. 14d). Thus, denoting by $w(n)$ the maximum width of a drawing constructed by the algorithm, $w(n) = 1 + w(n_1) + w(n_2)$ holds, where n_1 and n_2 are the maximum number of nodes in a left subtree of P and in a right subtree of P, respectively [assuming that $w(n)$ is monotone with n]. Thus, depending on the way in which P is chosen, different upper bounds on the asymptotic behavior of $w(n)$ can be achieved. Chan proves that P can be chosen so that $w(n) = n^{0.48}$. Such a bound is at the base of the best upper bound for constructing straight-line drawings of outerplanar graphs (see [42] and Sect. 3.4). An improvement on the following problem would be likely to improve the area upper bound on straight-line drawings of outerplanar graphs.

Open Problem 10. *Let $w(n)$ be the smallest function such that $w(1) = 1$ and such that, for every n-node ordered binary tree T, there exists a path P in T such $w(n) \leq 1 + w(n_1) + w(n_2)$, where every left subtree of P has at most n_1 nodes and every right subtree of P has at most n_2 nodes. What is the asymptotic behavior of $w(n)$?*

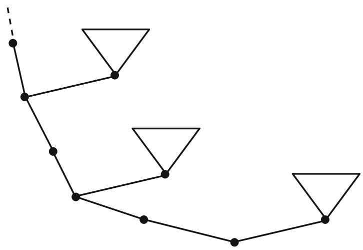


Fig. 15 A binary tree requiring $\Omega(n \log n)$ area in any straight-line upward orthogonal drawing. The tree is composed of a path P and of complete binary trees with size $n^{\alpha/2}$, where α is some constant greater than 0, attached to the ith node of P, for each i multiple of $n^{\alpha/2}$

It is easy to observe an $\Omega(\log n)$ lower bound for $w(n)$. We believe that in fact $w(n) = \Omega(2^{\sqrt{\log n}})$, but it is not clear to us whether the same bound can be achieved from above.

Turning the attention back to straight-line strictly upward order-preserving drawings, the following problem remains open.

Open Problem 11. *Close the gap between the $O(n4^{\sqrt{2\log n}})$ upper bound and the $\Omega(n \log n)$ lower bound for the area requirements of straight-line strictly upward order-preserving drawings of trees.*

Concerning *straight-line orthogonal drawings*, Chan et al. in [23] and Shin et al. in [112] have independently shown that $O(n \log\log n)$ area suffices for binary trees. Both algorithms are based on nice inductive geometric constructions and on nontrivial tree decompositions. Frati proved in [66] that every ternary tree admits a straight-line orthogonal drawing in $O(n^{1.631})$ area. The following problems are still open.

Open Problem 12. *Close the gap between the $O(n \log\log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line orthogonal drawings of binary trees.*

Open Problem 13. *Close the gap between the $O(n^{1.631})$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line orthogonal drawings of ternary trees.*

Concerning *straight-line upward orthogonal drawings*, Crescenzi et al. [34] and Chan et al. in [23] have shown that $O(n \log n)$ area suffices for binary trees. Such an area bound is worst-case optimal, as proved in [23]. The tree providing the lower bound, shown in Fig. 15, consists of a path to which some complete binary trees are attached.

Concerning *straight-line order-preserving orthogonal drawings*, $O(n^{1.5})$ and $O(n^2)$ area upper bounds are known [66] for binary and ternary trees, respectively. Once again, such algorithms are based on simple inductive geometric constructions.

While the bound for ternary trees is tight, no superlinear lower bound is known for straight-line order-preserving orthogonal drawings of binary trees. Hence, the following is open.

Open Problem 14. *Close the gap between the $O(n^{1.5})$ upper bound and the $\Omega(n)$ lower bound for the area requirements of straight-line order-preserving orthogonal drawings of binary trees.*

4 Polyline Drawings

In this section, we discuss algorithms and bounds for constructing small-area planar polyline drawings of planar graphs and their subclasses. In Sect. 4.1 we deal with general planar graphs, in Sect. 4.2 we deal with series-parallel and outerplanar graphs, and in Sect. 4.3 we deal with trees. Table 3 summarizes the best-known area bounds for polyline planar drawings of planar graphs and their subclasses. Observe that the lower bound of the table referring to general planar graphs has been obtained considering plane graphs.

4.1 General Planar Graphs

Every n-vertex plane graph admits a planar polyline drawing on a grid with $O(n^2)$ area. In fact, this has been known since the beginning of the 1980s [123]. Tamassia and Tollis introduced in [116] a technique that later became pretty much a standard for constructing planar polyline drawings. Namely, the authors showed that a polyline drawing Γ of a plane graph G can be easily obtained from a visibility representation R of G; moreover, Γ and R have asymptotically the same area. In order to obtain a visibility representation R of G, Tamassia and Tollis design a very nice algorithm (an application is shown in Fig. 16). The algorithm assumes that G is biconnected (if it is not, it suffices to augment G to biconnected by inserting dummy edges, apply the algorithm, and then remove the inserted dummy edges to obtain a visibility representation of G). The algorithm consists of the following steps: (1) Consider an orientation of G induced by an *st-numbering* of G, that is, a bijective

Table 3 A table summarizing the area requirements for polyline planar drawings of several classes of planar graphs

	Upper bound	Refs.	Lower bound	Refs.
General planar graphs	$\frac{4(n-1)^2}{9}$	[18]	$\frac{4(n-1)^2}{9}$	[37]
Series-parallel graphs	$O(n^{1.5})$	[13]	$\Omega(n2^{\sqrt{\log n}})$	[67]
Outerplanar graphs	$O(n \log n)$	[12, 13]	$\Omega(n)$	Trivial
Trees	$O(n \log n)$	[34]	$\Omega(n)$	Trivial

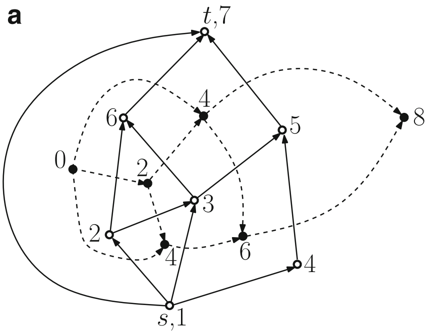


Fig. 16 An illustration for the algorithm of Tamassia and Tollis [116]. (**a**) *White circles* and *solid edges* represent G. *Black circles* and *dashed edges* represent G^*. An st-numbering of G (and the corresponding orientation) is shown. An orientation of G^* and the number $2\psi(f)$ for each each face f of G is shown. (**b**) A visibility representation of G

mapping $\phi : V(G) \rightarrow \{1,\ldots,n\}$ such that, for a given edge (s,t) incident to the outer face of G, $\phi(s) = 1$, $\phi(t) = n$, and for each $u \in V(G)$ with $u \neq s,t$, there exist two neighbors of u, say v and w, such that $\phi(v) < \phi(u) < \phi(w)$; (2) consider the orientation of the dual graph G^* of G induced by the orientation of G; (3) the y-coordinate of each vertex segment u is given by $\phi(u)$; (4) the y-coordinates of the endpoints of each edge segment (u,v) are given by $\phi(u)$ and $\phi(v)$; (5) the x-coordinate of edge segment (s,t) is set equal to -1; (6) the x-coordinate of each edge segment (u,v) is chosen to be any number strictly between $2\psi(f)$ and $2\psi(g)$, where f and g are the faces adjacent to (u,v) in G and $\psi(f)$ denotes the length of the longest path from the source to f in G^*; (7) finally, the x-coordinates of the endpoints of each vertex segment u are set equal to the smallest and largest x-coordinates of its incident edges.

Since the algorithm of Tamassia and Tollis, a large number of algorithms have been proposed to construct polyline drawings of planar graphs (see, e.g., [25, 79, 81, 124, 125]), proposing several tradeoffs among the area requirements, number of bends, and angular resolution. Here we briefly discuss an algorithm proposed by Bonichon et al. in [18], the first one to achieve optimal area, namely, $\frac{4(n-1)^2}{9}$. The algorithm consists of two steps. In the first one, a deep study of Schnyder realizers (see [110] and Sect. 3.1 for the definition of Schnyder realizers) leads to the definition of a *weak-stratification* of a realizer. Namely, given a realizer (T_0, T_1, T_2) of a triangulation G, a weak stratification is a layering L of the vertices of G such that T_0 (which is rooted at the vertex incident to the outer face of G) is upward, while T_1 and T_2 (which are rooted at the vertices incident to the outer face of G) are downward and some further conditions are satisfied. Each vertex will get a y-coordinate, which is equal to its layer in the weak stratification. In the second step, x-coordinates for vertices and bends are computed. The conditions of the weak stratification ensure that a planar drawing can in fact be obtained.

4.2 Series-Parallel and Outerplanar Graphs

Biedl proved in [13] that every series-parallel graph admits a polyline drawing with $O(n^{1.5})$ area and a polyline drawing with $O(fn\log n)$ area, where f is the fan-out of the series-parallel graph. In particular, since outerplanar graphs are series-parallel graphs with fan-out 2, the last result implies that outerplanar graphs admit polyline drawings with $O(n\log n)$ area. Biedl's algorithm constructs a visibility representation R of the input graph G with $O(n^{1.5})$ area; a polyline drawing Γ with asymptotically the same area of R can then be easily obtained from R. In order to construct a visibility representation R of the input graph G, Biedl relies on a strong inductive hypothesis, namely, that a small-area visibility representation R of G can be constructed with the further constraint that the poles s and t of G are placed at the top right corner and at the bottom right corner of the representation. Figures 17a, b show how this is accomplished in the base case. The parallel case is also pretty simple, as the visibility representations of the components of G are just placed one beside the other (as in Figs. 17c, d). The series case is much more involved. Namely, assuming w.l.o.g. that G is the series of two components H_1 and H_2, where H_1 has poles s and x and H_2 has poles x and t, and assuming w.l.o.g. that H_2 has more vertices than H_1, then if H_2 is the parallel composition of a "small" number of components, the composition shown in Fig. 17e, f is applied, while if H_2 is the parallel composition of a "large" number of components, the composition shown in Fig. 17g, h is applied. The rough idea behind these constructions is that if H_2 is the parallel composition of a small number of components, then a vertical unit can be spent for each of them without increasing the height of the drawing much. On the other hand, if H_2 is the parallel composition of a large number of components, then lots of such components have few vertices; hence, two of them can be placed one above the other without increasing the height of the drawing much.

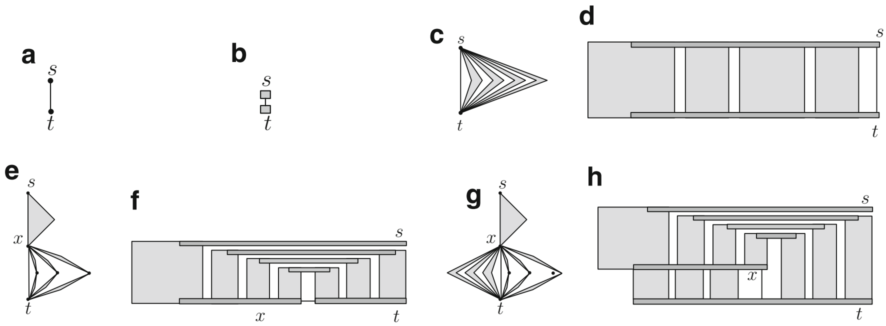


Fig. 17 Biedl's algorithm for constructing visibility representations of series-parallel graphs. (**a**, **b**) The base case. (**c**, **d**) The parallel case. (**e**, **h**) The series case

Table 4 Summary of the best-known area bounds for polyline drawings of trees

	Ord. Pres.	Upw.	Str. Upw.	Orth.	Upper bound	Refs.	Lower bound	Refs.
Binary					$O(n)$	[70]	$\Omega(n)$	Trivial
Binary	✓				$O(n \log \log n)$	[71]	$\Omega(n)$	Trivial
Binary		✓			$O(n)$	[70]	$\Omega(n)$	Trivial
Binary			✓		$O(n \log n)$	[34]	$\Omega(n \log n)$	[34]
Binary	✓		✓		$O(n \log n)$	[70]	$\Omega(n \log n)$	[34]
Binary				✓	$O(n)$	[121]	$\Omega(n)$	Trivial
Binary		✓		✓	$O(n \log \log n)$	[70]	$\Omega(n \log \log n)$	[70]
Binary	✓			✓	$O(n)$	[53]	$\Omega(n)$	Trivial
Binary	✓	✓		✓	$O(n \log n)$	[92]	$\Omega(n \log n)$	[70]
Ternary				✓	$O(n)$	[121]	$\Omega(n)$	Trivial
Ternary		✓		✓	$O(n \log n)$	[92]	$\Omega(n \log n)$	[92]
Ternary	✓			✓	$O(n)$	[53]	$\Omega(n)$	Trivial
Ternary	✓	✓		✓	$O(n \log n)$	[92]	$\Omega(n \log n)$	[70]
General					$O(n \log n)$	[34]	$\Omega(n)$	Trivial
General	✓				$O(n \log n)$	[71]	$\Omega(n)$	Trivial
General		✓			$O(n \log n)$	[34]	$\Omega(n)$	Trivial
General			✓		$O(n \log n)$	[34]	$\Omega(n \log n)$	[34]
General	✓		✓		$O(n 4^{\sqrt{2 \log n}})$	[24]	$\Omega(n \log n)$	[34]

"Ord. Pres.," "Upw.," "Str. Upw.," and "Orth." stand for order-preserving, upward, strictly upward, and orthogonal, respectively.

The following problems remain open.

Open Problem 15. *Close the gap between the* $O(n^{1.5})$ *upper bound and the* $\Omega(n 2^{\sqrt{\log n}})$ *lower bound for the area requirements of polyline drawings of series-parallel graphs.*

Open Problem 16. *Close the gap between the* $O(n \log n)$ *upper bound and the* $\Omega(n)$ *lower bound for the area requirements of polyline drawings of outerplanar graphs.*

4.3 Trees

No algorithms are known that exploit the possibility of bending the edges of a tree to get area bounds better than the corresponding ones shown for straight-line drawings.

Open Problem 17. *Close the gap between the* $O(n \log n)$ *upper bound and the* $\Omega(n)$ *lower bound for the area requirements of polyline drawings of trees.*

However, better bounds can be achieved for polyline drawings satisfying further constraints. Table 4 summarizes the best-known area bounds for various kinds of polyline drawings of trees.

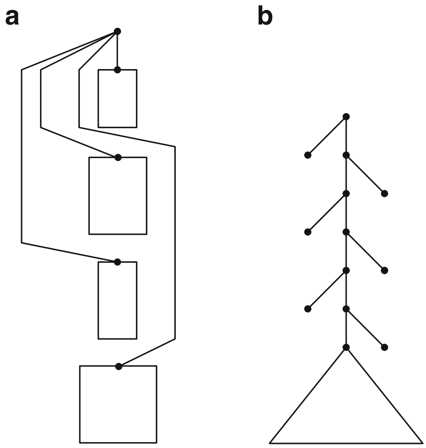


Fig. 18 (**a**) The construction of Garg et al. [70] to obtain $O(n \log n)$ area polyline order-preserving strictly upward drawings of bounded-degree trees. (**b**) A tree requiring $\Omega(n \log n)$ area in any upward order-preserving drawing. The *triangle* represents a complete binary tree with $n/3$ nodes

Concerning *polyline upward drawings*, a linear area bound is known, due to Garg *et al.* [70], for all trees whose degree is $O(n^{\delta})$, where δ is *any* constant less than 1. The algorithm of Garg et al. first constructs a layering $\gamma(T)$ of the input tree T; in $\gamma(T)$ each node u is assigned a layer smaller than or equal to the layer of the leftmost child of u and smaller than the layer of any other child of u; second, the authors show that $\gamma(T)$ can be converted into an upward polyline drawing whose height is the number of layers and whose width is the maximum *width of a layer*, that is, the number of nodes of the layer plus the number of edges crossing the layer; third, the authors show how to construct a layering of every tree whose degree is $O(n^{\delta})$ so that the number of layers times the maximum width of a layer is $O(n)$. No upper bound better than $O(n \log n)$ (from the results on straight-line drawings; see [34] and Sect. 3.5) and no superlinear lower bound is known for trees with unbounded degree.

Open Problem 18. *Close the gap between the* $O(n \log n)$ *upper bound and the* $\Omega(n)$ *lower bound for the area requirements of polyline upward drawings of trees.*

Concerning *polyline order-preserving strictly upward drawings*, Garg et al. [70] show a simple algorithm to achieve $O(n \log n)$ area for bounded-degree trees. The algorithm, whose construction is shown in Fig. 18a, consists of stacking inductively constructed drawings of the subtrees of the root of the input tree T in such a way that the tree with the greatest number of nodes is the bottommost in the drawing. The edges connecting the root to its subtrees are then routed beside the subtrees. The $O(n \log n)$ area upper bound is tight. Namely, there exist binary trees requiring $\Omega(n \log n)$ area in any strictly-upward order-preserving drawing [34] and binary trees requiring $\Omega(n \log n)$ area in any (even nonstrictly) upward order-preserving drawing [70]. The lower-bound tree of Garg et al. is shown in Fig. 18b. As far as we know, no area bounds better than the ones for straight-line drawings have been proved for general trees; hence, the following are open.

Open Problem 19. *Close the gap between the $O(n\log n)$ upper bound and the $\Omega(n)$ lower bound for the area requirements of polyline order-preserving drawings of trees.*

Open Problem 20. *Close the gap between the $O(n4^{\sqrt{\log n}})$ upper bound and the $\Omega(n\log n)$ lower bound for the area requirements of polyline order-preserving strictly upward drawings of trees.*

Concerning *orthogonal drawings*, Valiant proved in [121] that every n-node ternary tree (and every n-node binary tree) admits a $\Theta(n)$ area orthogonal drawing. Such a result was strengthened by Dolev and Trickey in [53], who proved that ternary trees (and binary trees) admit $\Theta(n)$ area order-preserving orthogonal drawings. The technique of Valiant is based on the use of separator edges (see [121] and Sect. 3.5). The result of Dolev and Trickey is a consequence of a more general result on the construction of linear-area embeddings of degree-4 outerplanar graphs.

Concerning *orthogonal upward drawings*, an $O(n\log\log n)$ area bound for binary trees was proved by Garg et al. in [70]. The algorithm has several ingredients. (1) A simple algorithm is shown to construct orthogonal upward drawings in $O(n\log n)$ area; such drawings exhibit the further property that no vertical line through a node of degree at most 2 intersects the drawing below such a node. (2) The *separator tree* S of the input tree T is constructed; such a tree represents the recursive decomposition of a tree via separator edges; namely, S is a binary tree that is recursively constructed as follows: The root r of S is associated with tree T and with a separator edge of T that splits T into subtrees T_1 and T_2; the subtrees of r are the separator trees associated with T_1 and T_2; observe that the leaves of S are the nodes of T. (3) A *truncated separator tree* S' is obtained from S by removing all the nodes of S associated with subtrees of T with less than $\log n$ nodes. (4) Drawings of the subtrees of T associated with the leaves of S' are constructed via the $O(n\log n)$ area algorithm. (5) Such drawings are stacked one on top of the other and the separator edges connecting them are routed (see Fig. 19a). The authors prove that the constructed drawings have $O(\frac{n\log\log n}{\log n})$ height and $O(\log n)$ width, thus obtaining the claimed upper bound. The same authors also proved that the $O(n\log\log n)$ bound is tight, by exhibiting the class of trees shown in Fig. 19b. In [92] Kim showed that $\Theta(n\log n)$ area is an optimal bound for upward orthogonal drawings of ternary trees. The upper bound comes from a stronger result on orthogonal order-preserving upward drawings cited below, while the lower bound comes from the tree shown in Fig. 19c.

Concerning *orthogonal order-preserving upward drawings*, $\Theta(n\log n)$ is an optimal bound both for binary and ternary trees. In fact, Kim [92] proved the upper bound for ternary trees (such a bound can be immediately extended to binary trees). The simple construction of Kim is presented in Fig. 20. The lower bound comes directly from the results of Garg et al. on order-preserving upward (nonorthogonal) drawings [70].

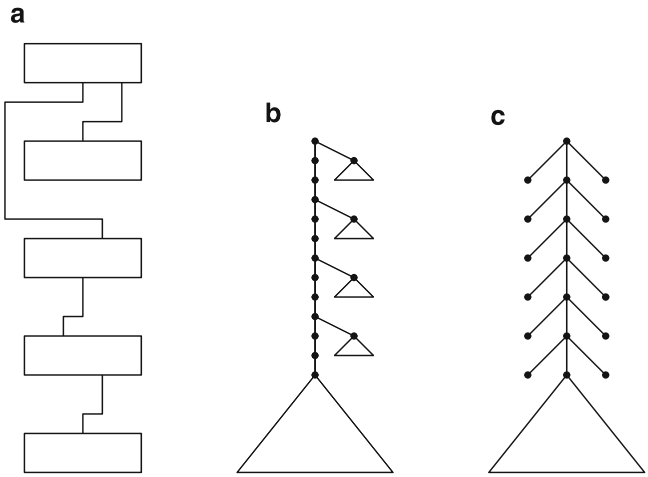


Fig. 19 (**a**) The construction of Garg et al. [70] to obtain $O(n \log \log n)$ area orthogonal upward drawings of binary trees. *Rectangles* represent drawings of small subtrees constructed via an $O(n \log n)$ area algorithm. (**b**) A binary tree requiring $\Omega(n \log \log n)$ area in any upward orthogonal drawing. The tree is composed of a chain with $n/3$ nodes, a complete binary tree with $n/3$ nodes (the *large triangle* in the figure), and $\frac{n}{3\sqrt{\log n}}$ subtrees (the *small triangles* in the figure) with $\sqrt{\log n}$ nodes rooted at the child of each $\sqrt{\log n}$th node of the chain. (**c**) A ternary tree requiring $\Omega(n \log n)$ area in any upward orthogonal drawing. The tree is composed of a chain with $n/4$ nodes, two other children for each node of the chain, and a complete binary tree with $n/4$ nodes (the *large triangle* in the figure)

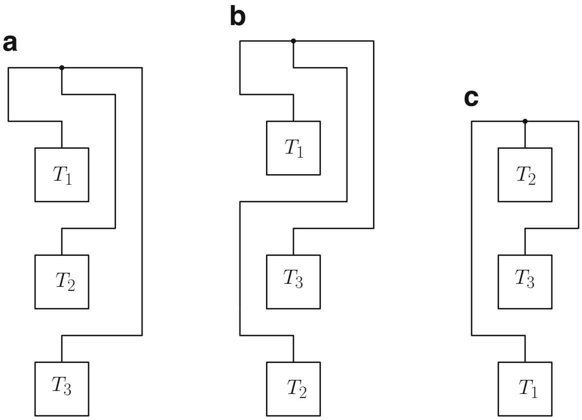


Fig. 20 An algorithm to construct $O(n \log n)$ area orthogonal order-preserving upward drawings of ternary trees. The figures illustrate the cases in which: (**a**) the right subtree has the greatest number of nodes; (**b**) the middle subtree has the greatest number of nodes; and (**c**) the left subtree has the greatest number of nodes

5 Upward Drawings

In this section, we discuss algorithms and bounds for constructing small-area planar straight-line/polyline upward drawings of upward planar directed acyclic graphs.

Table 5 A table summarizing the area requirements for straight-line upward planar drawings of upward planar DAGs; b and c denote constants greater than 1

	Upper bound	Refs.	Lower bound	Refs.
General upward planar DAGs	$O(c^n)$	[75]	$\Omega(b^n)$	[48]
Fixed-embedding series-parallel DAGs	$O(c^n)$	[75]	$\Omega(b^n)$	[10]
Series-parallel DAGs	$O(n^2)$	[10]	$\Omega(n^2)$	Trivial
Bipartite DAGs	$O(c^n)$	[75]	$\Omega(b^n)$	[65]
Fixed-embedding directed trees	$O(c^n)$	[75]	$\Omega(b^n)$	[65]
Directed trees	$O(n \log n)$	[65]	$\Omega(n \log n)$	[65]

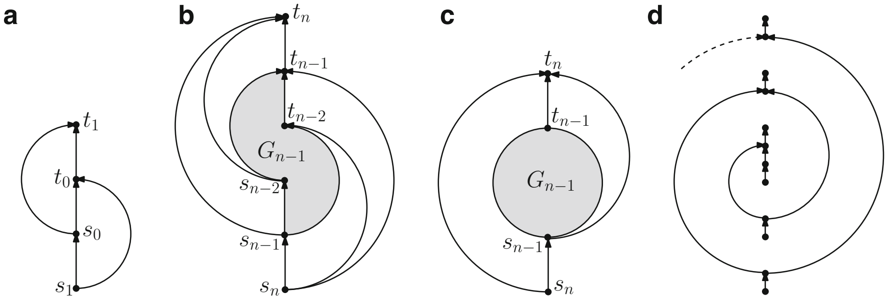


Fig. 21 (**a**, **b**) Inductive construction of a class G_n of upward planar DAGs requiring exponential area in any planar straight-line upward drawing. (**c**) Inductive construction of a class of series-parallel DAGs requiring exponential area in any planar straight-line upward drawing respecting a fixed embedding. (**d**) A class of directed trees requiring exponential area in any planar straight-line upward drawing respecting a fixed embedding

Table 5 summarizes the best-known area bounds for straight-line upward planar drawings of upward planar DAGs and their subclasses.

It is known that testing the upward planarity of a DAG is an NP-complete problem if the DAG has a variable embedding [76], while it is polynomial-time solvable if the embedding of the DAG is fixed [11], if the underlying graph is an outerplanar graph [104], if the DAG has a single source [84], or if it is bipartite [45]. Di Battista and Tamassia [46] showed that a DAG is upward planar if and only if it is a subgraph of an st-planar DAG. Some families of DAGs are always upward planar, like the series-parallel DAGs and the directed trees.

Di Battista and Tamassia proved in [46] that every upward planar DAG admits an upward straight-line drawing. Such a result is achieved by means of an algorithm similar to Fáry's algorithm for constructing planar straight-line drawings of undirected planar graphs (see Sect. 3.1). However, while planar straight-line drawings of undirected planar graphs can be constructed in polynomial area, Di Battista et al. proved in [48] that there exist upward planar DAGs that require exponential area in any planar straight-line upward drawing. Such a result is achieved by considering the class G_n of DAGs whose inductive construction is shown in Figs. 21a, b and by using some geometric considerations to prove that the area of the smallest region containing an upward planar straight-line drawing of G_n is a constant number of times larger than the area of a region containing an upward planar straight-line drawing of G_{n-1}. The techniques introduced by Di Battista et al. in [48] to prove the exponential lower bound for the area requirements of upward planar straight-line drawings of upward planar DAGs have later been strengthened by Bertolazzi et al. in [10] and by Frati in [65] to prove, respectively, that there exist series-parallel DAGs with fixed embedding (see Fig. 21c) and there exist directed trees with fixed embedding (see Fig. 21d) requiring exponential area in any upward planar straight-line drawing. Similar lower-bound techniques have also been used to deal with straight-line drawings of clustered graphs (see Sect. 8).

On the positive side, area-efficient algorithms exist for constructing upward planar straight-line drawings for restricted classes of upward planar DAGs. Namely, Bertolazzi et al. in [10] have shown how to construct upward planar straight-line drawings of series-parallel DAGs in optimal $\Theta(n^2)$ area, and Frati [65] has shown how to construct upward planar straight-line drawings of directed trees in optimal $\Theta(n \log n)$ area. Both algorithms are based on the inductive construction of upward planar straight-line drawings satisfying some additional geometric constraints. We remark that for upward planar DAGs whose underlying graph is a series-parallel graph, neither an exponential lower bound nor a polynomial upper bound is known for the area requirements of straight-line upward planar drawings. Observe that testing upward planarity for this family of graphs can be done in polynomial time [50].

Open Problem 21. *What are the area requirements of straight-line upward planar drawings of upward planar DAGs whose underlying graph is a series-parallel graph?*

Algorithms have been provided to construct upward planar polyline drawings of upward planar DAGs. The first $\Theta(n^2)$ optimal area upper bound for such drawings has been established by Di Battista and Tamassia in [46]. Their algorithm consists of first constructing an upward visibility representation of the given upward planar DAG and then of turning such a representation into an upward polyline drawing. Such a technique has been discussed in Sect. 4.

6 Convex Drawings

In this section, we discuss algorithms and bounds for constructing small-area convex and strictly convex drawings of planar graphs. Table 6 summarizes the best-known area bounds for convex and strictly convex drawings of planar graphs.

Not every planar graph admits a convex drawing. Tutte [119, 120] proved that every triconnected planar graph G admits a strictly convex drawing in which its outer face is drawn as an arbitrary strictly convex polygon P. His algorithm consists of first drawing the outer face of G as P and then placing each vertex at the barycenter

Table 6 A table summarizing the area requirements for convex and strictly convex drawings of triconnected plane graphs

	Upper bound	Refs.	Lower bound	Refs.
Convex	$n^2 + O(n)$	[17, 28, 47, 111]	$\frac{4n^2}{9} - O(n)$	[37, 69, 100, 121]
Strictly convex	$O(n^4)$	[8]	$\Omega(n^3)$	[1, 7, 9, 106]

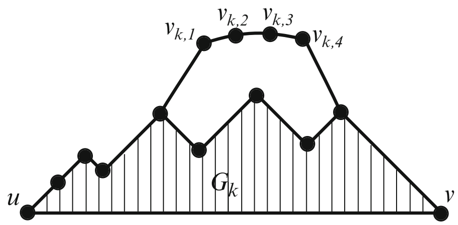


Fig. 22 An illustration of the canonical ordering of a triconnected plane graph

of the positions of its adjacent vertices. This results in a set of linear equations that always admits a unique solution.

Characterizations of the plane graphs admitting convex drawings were given by Tutte in [119, 120], by Thomassen in [117, 118], by Chiba et al. in [26], by Nishizeki and Chiba in [102], and by Di Battista et al. in [49]. Roughly speaking, the plane graphs admitting convex drawings are biconnected, their separation pairs are composed of vertices both incident to the outer face, and distinct separation pairs do not "nest." Chiba et al. presented in [26] a linear-time algorithm for testing whether a graph admits a convex drawing and producing a convex drawing if the graph allows for one. The area requirements of convex and strictly convex grid drawings have been widely studied, especially for triconnected plane graphs.

Convex grid drawings of triconnected plane graphs can be realized on a quadratic-size grid. This was first shown by Kant in [88]. In fact, Kant proved that such drawings can always be realized on a $(2n-4) \times (n-2)$ grid. The result is achieved by defining a stronger notion of canonical ordering of a plane graph (see Sect. 3.1). Such a strengthened canonical ordering allows us to construct every triconnected plane graph G starting from a cycle delimiting an internal face of G and repeatedly adding to the previously constructed biconnected graph G_k a vertex or a path in the outer face of G_k so that the newly formed graph G_{k+1} is also biconnected (see Fig. 22). Observe that this generalization of the canonical ordering allows us to deal with plane graphs containing nontriangular faces. Similarly to de Fraysseix et al.'s algorithm [37], Kant's algorithm exploits a canonical ordering of G to incrementally construct a convex drawing of G in which the outer face of the currently considered graph G_k is composed of segments whose slopes are $-45°$, or $0°$, or $45°$.

The bound of Kant was later improved down to $(n-2) \times (n-2)$ by Chrobak and Kant [28], and independently by Schnyder and Trotter [111]. The result of Chrobak and Kant again relies on a canonical ordering. On the other hand, the result of Schnyder and Trotter relies on a generalization of the Schnyder realizers (see

Sect. 3.1) in order to deal with triconnected plane graphs. Such an extension was independently shown by Di Battista et al. [47], who proved that every triconnected plane graph has a convex drawing on an $(f-2)\times(f-2)$ grid, where f is the number of faces of the graph. The best bound is currently, as far as we know, an $(n-2-\Delta)\times(n-2-\Delta)$ bound achieved by Bonichon, Felsner, and Mosbah in [17]. The bound is again achieved using Schnyder realizers. The parameter Δ is dependent on the Schnyder realizers and can vary among 0 and $\frac{n}{2}-2$. The following remains open.

Open Problem 22. *Close the gap between the* $(n-2-\Delta)\times(n-2-\Delta)$ *upper bound and the* $\frac{4n^2}{9}-O(n)$ *lower bound for the area requirements of convex drawings of triconnected plane graphs.*

Strictly convex drawings of triconnected plane graphs might require $\Omega(n^3)$ area. In fact, an n-vertex cycle needs $\Omega(n^3)$ area in any grid realization (see, e.g., [1,7,9]). The current best lower bound for the area requirements of a strictly convex polygon drawn on the grid, which has been proved by Rabinowitz in [106], is $\frac{n^3}{8\pi^2}$. The first polynomial upper bound for strictly convex drawings of triconnected plane graphs has been proved by Chrobak et al. in [27]. The authors showed that every triconnected plane graph admits a strictly convex drawing in an $O(n^3)\times O(n^3)$ grid. Their idea consists of first constructing a (non strictly) convex drawing of the input graph and of then perturbing the positions of the vertices in order to achieve strict convexity. A more elaborate technique relying on the same idea allowed Rote to achieve an $O(n^{7/3})\times O(n^{7/3})$ area upper bound in [109], which was further improved by Bárány and Rote to $O(n^2)\times O(n^2)$ and to $O(n)\times O(n^3)$ in [8]. The last ones are, as far as we know, the best-known upper bounds. One of the main differences between Chrobak et al.'s algorithm and Bárány and Rote's ones is that the former one constructs the intermediate nonstrictly convex drawing by making use of a canonical ordering of the graph, while the latter ones by making use of the Schnyder realizers. The following is, in our opinion, a very nice open problem.

Open Problem 23. *Close the gap between the* $O(n^4)$ *upper bound and the* $\Omega(n^3)$ *lower bound for the area requirements of strictly convex drawings of triconnected plane graphs.*

7 Proximity Drawings

In this section, we discuss algorithms and bounds for constructing small-area proximity drawings of planar graphs.

Characterizing the graphs that admit a proximity drawing, for a certain definition of proximity, is a difficult problem. For example, despite several research efforts (see, e.g., [51, 52, 95]), characterizing the graphs that admit a *realization* (a word that often substitutes *drawing* in the context of proximity graphs) as Delaunay triangulations is still an intriguing open problem. Dillencourt showed that every

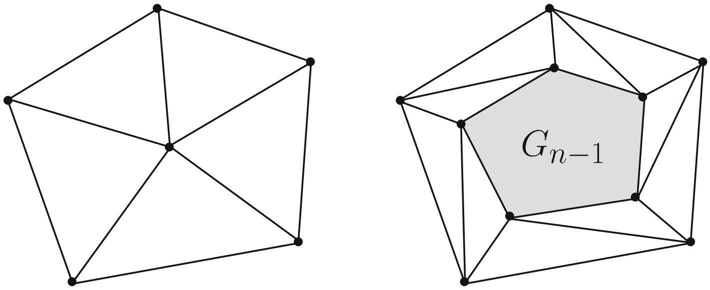


Fig. 23 Inductive construction of a class G_n of graphs requiring exponential area in any Gabriel drawing, in any weak Gabriel drawing, and in any β-drawing

maximal outerplanar graph can be realized as a Delaunay triangulation [51] and provided examples of small triangulations that cannot. The decision version of several realizability problems (that is, given a graph G and a definition of proximity, can G be realized as a proximity graph?) is $\mathcal{NP}$-hard. For example, Eades and Whitesides proved that deciding whether a tree can be realized as a minimum spanning tree is an $\mathcal{NP}$-hard problem [57], and that deciding whether a graph can be realized as a nearest-neighbor graph is an $\mathcal{NP}$-hard problem [56] as well. Both proofs rely on a mechanism for providing the hardness of graph-drawing problems, called a *logic engine*, which is interesting by itself. On the other hand, for several definitions of proximity graphs (such as Gabriel graphs and relative neighborhood graphs), the realizability problem is polynomial-time solvable for trees, as shown by Bose et al. [19]; further, Lubiw and Sleumer proved that maximal outerplanar graphs can be realized as relative neighborhood graphs and Gabriel graphs [98], a result later extended by Lenhart and Liotta to all biconnected outerplanar graphs [95]. For more results about proximity drawings, see [43, 96].

Most of the known algorithms to construct proximity drawings produce representations whose size increases exponentially with the number of vertices (see, e.g., [19, 44, 95, 98]). This seems to be unavoidable for most kinds of proximity drawings, although few exponential-area lower bounds are known. Liotta et al. [97] showed a class of graphs (whose inductive construction is shown in Fig. 23) requiring exponential area in any Gabriel drawing, in any weak Gabriel drawing, and in any β-drawing. Their proof is based on the observation that the circles whose diameters are the segments representing the edges incident to the outer face of G_n cannot contain any point in their interior. Consequently, the vertices of G_{n-1} are allowed only to be placed in a region whose area is a constant number of times smaller than the area of G_n. On the other hand, Penna and Vocca [105] showed algorithms to construct polynomial-area weak Gabriel drawings and weak β-drawings of binary and ternary trees.

Particular attention has been devoted to the area requirements of Euclidean minimum spanning trees. In their seminal paper on Euclidean minimum spanning trees, Monma and Suri [101] proved that any tree of maximum degree 5 admits a

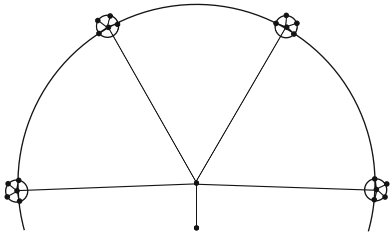


Fig. 24 An illustration of the algorithm of Monma and Suri to construct realizations of degree-5 trees as Euclidean minimum spanning trees

planar embedding as a Euclidean minimum spanning tree. Their algorithm, whose inductive construction is shown in Fig. 24, consists of placing the neighbors r_i of the root r of the tree on a circumference centered at r, of placing the neighbors of r_i on a much smaller circumference centered at r_i, and so on. Monma and Suri [101] proved that the area of the realizations constructed by their algorithm is $2^{\Omega(n^2)}$ and conjectured that exponential area is sometimes required to construct realizations of degree-5 trees as Euclidean minimum spanning trees. Kaufmann [90] and Frati and Kaufmann [68] showed how to construct polynomial-area realizations of degree-4 trees as Euclidean minimum spanning trees. Their technique consists of using a decomposition of the input tree T (similar to the ones presented in Sects. 3.4 and 3.5) in which a path P is selected such that every subtree of P has at most $n/2$ nodes. Euclidean minimum spanning tree realizations of such subtrees are then inductively constructed and placed together with a drawing of P to get a drawing of T. Suitable angles and lengths for the edges in P have to be chosen to ensure that the resulting drawing is a Euclidean minimum spanning tree realization of T. The sketched geometric construction is shown in Fig. 25. Very recently, Angelini et al. proved in [3] that in fact there exist degree-5 trees requiring exponential area in any realization as a Euclidean minimum spanning tree. The tree T^* exhibited by Angelini et al., which is shown in Fig. 26, consists of a degree-5 complete tree T_c with a constant number of vertices and of a set of degree-5 caterpillars, each one attached to a distinct leaf of T_c. The complete tree T_c forces the angles incident to an end vertex of the backbone of at least one of the caterpillars to be very small, that is, between 60° and 61°. Using this as a starting point, Angelini et al. prove that each angle incident to a vertex of the caterpillar is either very small, that is, between 60° and 61°, or is very large, that is, between 89.5° and 90.5°. As a consequence, the lengths of the edges of the backbone of the caterpillar decrease exponentially along the caterpillar, thus obtaining the area bound. There is still some distance between the best-known lower and upper bounds; hence, the following is open.

Open Problem 24. *Close the gap between the $2^{O(n^2)}$ upper bound and the $2^{\Omega(n)}$ lower bound for the area requirements of Euclidean minimum spanning tree realizations.*

Greedy drawings are a kind of proximity drawings that have recently attracted lot of attention due to their application to network routing. Namely, consider a network in which each node a that has to send a packet to some node b forwards the packet

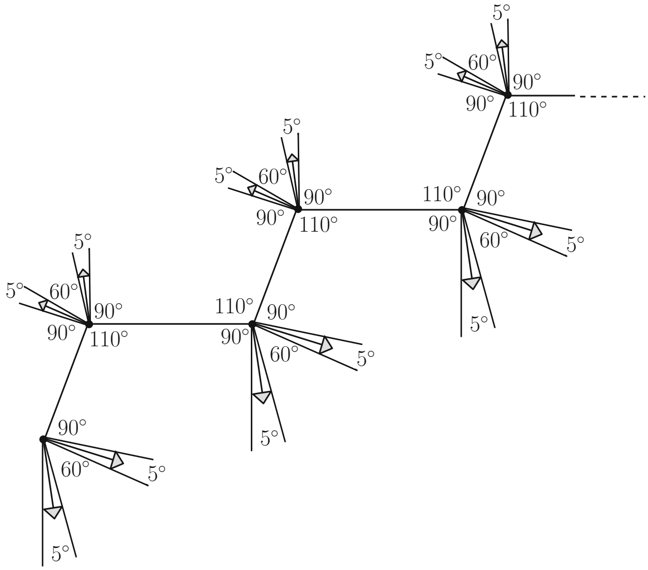


Fig. 25 An illustration of the algorithm of Frati and Kaufmann to construct polynomial-area realizations of degree-4 trees as Euclidean minimum spanning tree realizations

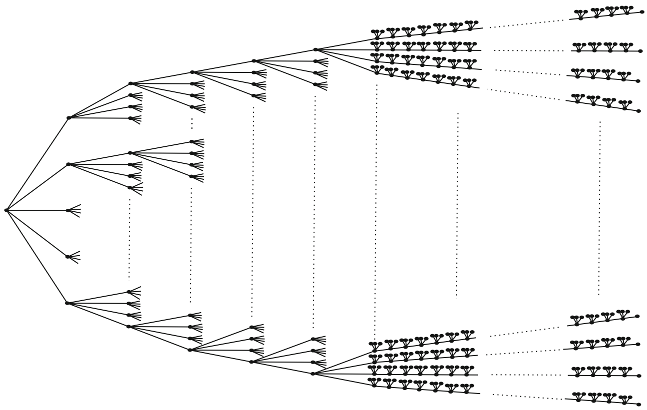


Fig. 26 A tree T^* requiring $2^{\Omega(n)}$ area in any Euclidean minimum spanning tree realization

to any node c that is closer to b than a itself. If the position of any node u is not its real geographic location, but rather the pair of coordinates of u in a drawing Γ of the network, it is easy to see that routing protocol never gets stuck if and only if Γ is a greedy drawing. Greedy drawings were introduced by Rao et al. in [107]. A lot of attention has been devoted to a conjecture of [103] stating that every triconnected planar graph has a greedy drawing. Dhandapani verified the conjecture for triangulations in [38], and later Leighton and Moitra [94] and independently

Angelini et al. [4] completely settled the conjecture in the positive. The approach of Leighton and Moitra (the one of Angelini et al. is amazingly similar) consists of finding a certain subgraph of the input triconnected planar graph, called a *cactus graph*, and of constructing a drawing of the cactus by induction. Greedy drawings have been proved to exist for every graph if the coordinates are chosen in the hyperbolic plane [93]. Research efforts have also been devoted to construct greedy drawings in a small area. More precisely, because of the routing applications, attention has been devoted to the possibility of encoding the coordinates of a greedy drawing with a small number of bits. When this is possible, the drawing is called *succinct*. Eppstein and Goodrich [58] and Goodrich and Strash [78] showed how to modify the algorithm of Kleinberg [93] and the algorithm of Leighton and Moitra [94], respectively, in order to construct drawings in which the vertex coordinates are represented by a logarithmic number of bits. On the other hand, Angelini et al. [2] proved that there exist trees requiring exponential area in any greedy drawing (or equivalently requiring a polynomial number of bits to represent their Cartesian coordinates in the Euclidean plane). The following is open, however.

Open Problem 25. *Is it possible to construct greedy drawings of triconnected planar graphs in the Euclidean plane in polynomial area?*

Partially positive results on the mentioned open problem were achieved by He and Zhang, who proved in [83] that succinct convex *weekly greedy* drawings exist for all triconnected planar graphs, where ”weakly greedy” means that the distance between two vertices u and v in the drawing is not the usual Euclidean distance $D(u,v)$ but a function $H(u,v)$ such that $D(u,v) \leq H(u,v) \leq 2\sqrt{2}D(u,v)$. On the other hand, Cao et al. proved in [22] that there exist triconnected planar graphs requiring exponential area in any *convex* greedy drawing in the Euclidean plane.

8 Clustered Graph Drawings

In this section, we discuss algorithms and bounds for constructing small-area c-planar drawings of clustered graphs. Table 7 summarizes the best-known area bounds for c-planar straight-line drawings of clustered graphs.

Given a clustered graph, testing whether it admits a c-planar drawing is a problem of unknown complexity and has perhaps been the most studied problem in the graph drawing community during the last 10 years [6, 30–33, 35, 41, 61, 63, 77, 80, 85–87].

Table 7 A table summarizing the area requirements for c-planar straight-line drawings of clustered graphs in which clusters are convex regions; b and c denote constants greater than 1

	Upper bound	Refs.	Lower bound	Refs.
Clustered graphs	$O(c^n)$	[5, 54]	$\Omega(b^n)$	[62]
c-Connected trees	$O(n^2)$	[39]	$\Omega(n^2)$	[39]
Non-c-connected trees	$O(c^n)$	[5, 54]	$\Omega(b^n)$	[39]

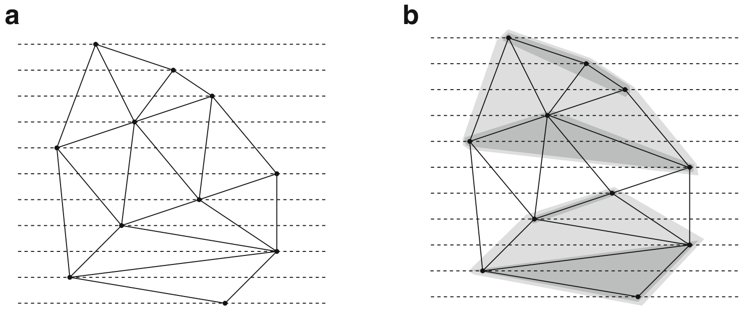


Fig. 27 (**a**) A planar straight-line drawing of a hierarchical graph H. Graph H is obtained from a clustered graph C by assigning consecutive layers to vertices of the same cluster. (**b**) A straight-line c-planar drawing of C

Suppose that a c-planar clustered graph C is given together with a c-planar embedding. How can the graph be drawn? Such a problem has been intensively studied in the literature, and a number of papers have been presented for constructing c-planar drawings of c-planar clustered graphs within many drawing conventions.

Eades et al. show in [54] an algorithm for constructing c-planar straight-line drawings of c-planar clustered graphs in which each cluster is drawn as a convex region. Such a result is achieved by first studying how to construct planar straight-line drawings of hierarchical graphs. A *hierarchical graph* is a graph such that each vertex v is assigned a number $y(v)$, called the *layer* of v; a drawing of a hierarchical graph has to place each vertex v on the horizontal line $y = y(v)$. Eades et al. show an inductive algorithm to construct a planar straight-line drawing of any hierarchical planar graph. Second, Eades et al. show how to turn a c-planar clustered graph C into a hierarchical graph H such that, for each cluster μ in C, all the vertices in μ appear in consecutive layers of the hierarchy. This implies that, once a planar straight-line drawing of H has been constructed, as in Fig. 27a, each cluster μ can be drawn as a region surrounding the convex hull of the vertices in μ, resulting in a straight-line c-planar drawing of C in which each cluster is drawn as a convex region, as in Fig. 27b.

Angelini et al., improving upon the described result of Eades et al. in [54] and answering a question posed in [54], show in [5] an algorithm for constructing a *straight-line rectangular drawing* of any clustered graph C, that is, a c-planar straight-line drawings of C in which each cluster is drawn as an axis-parallel rectangle (more in general, the algorithm of Angelini et al. constructs straight-line c-planar drawings in which each cluster is an arbitrary convex shape). The algorithm of Angelini et al. is reminiscent of Fáry's algorithm (see [59] and Sect. 3.1). Namely, the algorithm turns a clustered graph C into a smaller clustered graph C' by removing a cluster, or splitting C in correspondence of a separating 3-cycle, or contracting an edge of C. A straight-line rectangular drawing of C' can then be inductively constructed and easily augmented to a straight-line rectangular drawing of C. When none of the inductive cases applies, the clustered graph is an

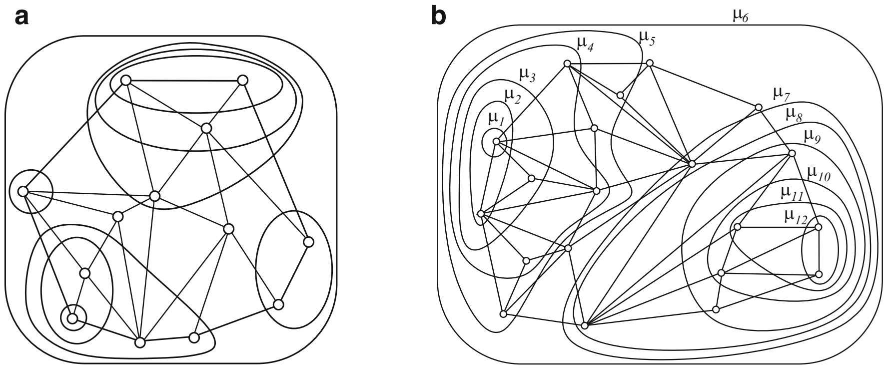


Fig. 28 (**a**) An outerclustered graph. (**b**) A linearly ordered outerclustered graph. Any two consecutive clusters in the sequence $\mu_1, \ldots, \mu_{12}$ are the parent of each other

outerclustered graph; that is, every cluster contains a vertex incident to the outer face (see Fig. 28a). In order to draw an outerclustered graph C, Angelini et al. show how to split C into three *linearly ordered outerclustered graphs*, which are outerclustered graphs such that the graph induced by the "direct containment" relationship among clusters is a path (see Fig. 28b), where a cluster μ *directly contains* a cluster ν if μ contains ν and μ contains no cluster ρ containing ν. Moreover, they show how to combine the drawings of such graphs to get a straight-line rectangular drawing of C. Finally, Angelini et al. show an inductive algorithm for constructing a straight-line rectangular drawing of any linearly ordered outerclustered graphs C. Such an algorithm finds a subgraph of C (a path plus an edge) that splits G into smaller linearly ordered outerclustered graphs, inductively draws such subgraphs, and combines their drawings to get a straight-line rectangular drawing of C.

Both the algorithm of Eades et al. and the algorithm of Angelini et al. construct drawings requiring, in general, exponential area. However, Feng et al. proved in [62] that there exists a clustered graph C requiring exponential area in any straight-line c-planar drawing in which the clusters are represented by convex regions. The proof of such a lower bound is strongly based on the proof of Di Battista et al. that there exist directed graphs requiring exponential area in any upward straight-line drawing (see [48] and Sect. 5). Eades et al. showed in [55] how to construct $O(n^2)$ area c-planar orthogonal drawings of clustered graphs with maximum degree 4; the authors first construct a visibility representation of the given clustered graph and then turn such a representation into an orthogonal drawing. Di Battista et al. [39] show algorithms for drawing clustered trees in a small area. In particular, they show an inductive algorithm to construct straight-line rectangular drawings of c-connected clustered trees in $O(n^2)$ area; however, they prove that there exist non-c-connected trees requiring exponential area in any straight-line drawing in which the clusters are represented by convex regions, again using the tools designed by Di Battista et al. in [48]. The following problem has been left open by Di Battista et al. [48].

Open Problem 26. *What are the area requirements of order-preserving straight-line c-planar drawings of clustered trees in which clusters are represented by convex regions?*

Acknowledgements This work is partially supported by the Italian Ministry of Research, Projects AlgoDEEP no. 2008TFBWL4 and FIRB "Advanced tracking system in intermodal freight transportation," no. RBIP06BZW8; by the Swiss National Science Foundation, Grant no. 200021-125287/1; and by the Centre Interfacultaire Bernoulli (CIB) of EPFL.

References

1. G.E. Andrews, A lower bound for the volumes of strictly convex bodies with many boundary points. Trans. Am. Math. Soc. **106**, 270–279 (1963)
2. P. Angelini, G. Di Battista, F. Frati, Succinct greedy drawings do not always exist, in *Graph Drawing (GD'09)*, ed. by D. Eppstein, E.R. Gansner. LNCS, Springer, vol. 5849 (2010) pp. 171–182
3. P. Angelini, T. Bruckdorfer, M. Chiesa, F. Frati, M. Kaufmann, C. Squarcella, On the area requirements of Euclidean minimum spanning trees (2011), manuscript
4. P. Angelini, F. Frati, L. Grilli, An algorithm to construct greedy drawings of triangulations. J. Graph Algorithm. Appl. **14**(1), 19–51 (2010)
5. P. Angelini, F. Frati, M. Kaufmann, Straight-line rectangular drawings of clustered graphs. Discr. Comput. Geom. **45**(1), 88–140 (2011)
6. P. Angelini, F. Frati, M. Patrignani, Splitting clusters to get c-planarity, in *Graph Drawing (GD'09)*, ed. by D. Eppstein, E.R. Gansner. LNCS, Springer, vol. 5849 (2010), pp. 57–68
7. I. Bárány, J. Pach, On the number of convex lattice polygons. Comb. Prob. Comput. **1**, 295–302 (1992)
8. I. Bárány, G. Rote, Strictly convex drawings of planar graphs. Doc. Math. **11**, 369–391 (2006)
9. I. Bárány, N. Tokushige, The minimum area of convex lattice n-gons. Combinatorica **24**(2), 171–185 (2004)
10. P. Bertolazzi, R.F. Cohen, G. Di Battista, R. Tamassia, I.G. Tollis, How to draw a series-parallel digraph. Int. J. Comput. Geom. Appl. **4**(4), 385–402 (1994)
11. P. Bertolazzi, G. Di Battista, G. Liotta, C. Mannino, Upward drawings of triconnected digraphs. Algorithmica **12**(6), 476–497 (1994)
12. T.C. Biedl, Drawing outer-planar graphs in $O(n \log n)$ area, in *Graph Drawing (GD'02)*, ed. by M.T. Goodrich. LNCS, Springer, vol. 2528 (2002), pp. 54–65
13. T.C. Biedl, Small drawings of outerplanar graphs, series-parallel graphs, and other planar graphs. Discr. Comput. Geom. **45**(1), 141–160 (2011)
14. T.C. Biedl, F.J. Brandenburg, Drawing planar bipartite graphs with small area, in *Canadian Conference on Computational Geometry (CCCG'05)*, University of Windsor, Ontario, Canada, 2005, pp. 105–108
15. T.C. Biedl, T.M. Chan, A. López-Ortiz, Drawing $K_{2,n}$: a lower bound. Inform. Process. Lett. **85**(6), 303–305 (2003)
16. T.C. Biedl, G. Kant, M. Kaufmann, On triangulating planar graphs under the four-connectivity constraint. Algorithmica **19**(4), 427–446 (1997)
17. N. Bonichon, S. Felsner, M. Mosbah, Convex drawings of 3-connected plane graphs. Algorithmica **47**(4), 399–420 (2007)
18. N. Bonichon, B. Le Saëc, M. Mosbah, Wagner's theorem on realizers, in *Automata, Languages and Programming (ICALP'02)*, ed. by P. Widmayer, F. Triguero Ruiz, R.M. Bueno, M. Hennessy, S. Eidenbenz, R. Conejo. LNCS, Springer, vol. 2380 (2002), pp. 1043–1053

19. P. Bose, W. Lenhart, G. Liotta, Characterizing proximity trees. Algorithmica **16**(1), 83–110 (1996)
20. F.J. Brandenburg, Drawing planar graphs on $\frac{8}{9}n^2$ area. Electron. Notes Discr. Math. **31**, 37–40 (2008)
21. A. Brocot, Calcul des rouages par approximation, nouvelle methode. Rev. Chronom. **6**, 186–194 (1860)
22. L. Cao, A. Strelzoff, J.Z. Sun, On succinctness of geometric greedy routing in Euclidean plane, in *Pervasive Systems, Algorithms, and Networks (ISPAN'09)*, Kaohsiung, Taiwan, 2009, pp. 326–331
23. T.M. Chan, M.T. Goodrich, S. Rao Kosaraju, R. Tamassia, Optimizing area and aspect ratio in straight-line orthogonal tree drawings. Comput. Geom.: Theor. Appl. **23**(2), 153–162 (2002)
24. T.M. Chan, A near-linear area bound for drawing binary trees. Algorithmica **34**(1), 1–13 (2002)
25. C.C. Cheng, C.A. Duncan, M.T. Goodrich, S.G. Kobourov, Drawing planar graphs with circular arcs. Discr. Comput. Geom. **25**(3), 405–418 (2001)
26. N. Chiba, T. Yamanouchi, T. Nishizeki, Linear algorithms for convex drawings of planar graphs, in *Progress in Graph Theory*, ed. by J.A. Bondy, U.S.R. Murty (Academic Press, New York, 1984), pp. 153–173
27. M. Chrobak, M.T. Goodrich, R. Tamassia, Convex drawings of graphs in two and three dimensions (preliminary version), in *Symposium on Computational Geometry (SoCG'96)*, Philadelphia, Pennsylvania, USA, 1996, pp. 319–328
28. M. Chrobak, G. Kant, Convex grid drawings of 3-connected planar graphs. Int. J. Comput. Geom. Appl. Limerick, Ireland, **7**(3), 211–223 (1997)
29. M. Chrobak, T.H. Payne, A linear-time algorithm for drawing a planar graph on a grid. Inform. Process. Lett. **54**(4), 241–246 (1995)
30. S. Cornelsen, D. Wagner, Completely connected clustered graphs. J. Discr. Algorithm. **4**(2), 313–323 (2006)
31. P.F. Cortese, G. Di Battista, F. Frati, M. Patrignani, M. Pizzonia, *C*-planarity of *c*-connected clustered graphs. J. Graph Algorithm. Appl. Carleton University, Ottawa, Canada, **12**(2), 225–262 (2008)
32. P.F. Cortese, G. Di Battista, M. Patrignani, M. Pizzonia, Clustering cycles into cycles of clusters. J. Graph Algorithm. Appl. **9**(3), 391–413 (2005)
33. P.F. Cortese, G. Di Battista, M. Patrignani, M. Pizzonia, On embedding a cycle in a plane graph. Discr. Math. **309**(7), 1856–1869 (2009)
34. P. Crescenzi, G. Di Battista, A. Piperno, A note on optimal area algorithms for upward drawings of binary trees. Comput. Geom.: Theor. Appl. **2**, 187–200 (1992)
35. E. Dahlhaus, A linear time algorithm to recognize clustered graphs and its parallelization, in *Latin American Symposium on Theoretical Informatics (LATIN'98)*, ed. by C.L. Lucchesi, A.V. Moura. LNCS, Springer, vol. 1380 (1998), pp. 239–248
36. H. de Fraysseix, J. Pach, R. Pollack, Small sets supporting Fáry embeddings of planar graphs, in *Symposium on Theory of Computing (STOC'88)*, ed. by J. Simon, Chicago, Illinois, USA, (1988), pp. 426–433
37. H. de Fraysseix, J. Pach, R. Pollack, How to draw a planar graph on a grid. Combinatorica **10**(1), 41–51 (1990)
38. R. Dhandapani, Greedy drawings of triangulations. Discr. Comput. Geom. **43**(2), 375–392 (2010)
39. G. Di Battista, G. Drovandi, F. Frati, How to draw a clustered tree. J. Discr. Algorithm. **7**(4), 479–499 (2009)
40. G. Di Battista, P. Eades, R. Tamassia, I.G. Tollis, *Graph Drawing* (Prentice Hall, Upper Saddle River, NJ, 1999)
41. G. Di Battista, F. Frati, Efficient c-planarity testing for embedded flat clustered graphs with small faces. J. Graph Algorithm. Appl. **13**(3), 349–378 (2009)
42. G. Di Battista, F. Frati, Small area drawings of outerplanar graphs. Algorithmica **54**(1), 25–53 (2009)

43. G. Di Battista, W. Lenhart, G. Liotta, Proximity drawability: a survey, in *Graph Drawing (GD'94)*, ed. by R. Tamassia, I.G. Tollis. LNCS, Princeton, New Jersey, USA, vol. 894 (1994), pp. 328–339
44. G. Di Battista, G. Liotta, S. Whitesides, The strength of weak proximity. J. Discr. Algorithm. **4**(3), 384–400 (2006)
45. G. Di Battista, W.P. Liu, I. Rival, Bipartite graphs, upward drawings, and planarity. Inform. Process. Lett. **36**(6), 317–322 (1990)
46. G. Di Battista, R. Tamassia, Algorithms for plane representations of acyclic digraphs. Theor. Comput. Sci. **61**, 175–198 (1988)
47. G. Di Battista, R. Tamassia, L. Vismara, Output-sensitive reporting of disjoint paths. Algorithmica **23**(4), 302–340 (1999)
48. G. Di Battista, R. Tamassia, I.G. Tollis, Area requirement and symmetry display of planar upward drawings. Discr. Comput. Geom. **7**, 381–401 (1992)
49. G. Di Battista, R. Tamassia, L. Vismara, Incremental convex planarity testing. Inform. Comput. **169**(1), 94–126 (2001)
50. W. Didimo, F. Giordano, G. Liotta, Upward spirality and upward planarity testing, in *Graph Drawing (GD'05)*, ed. by P. Healy, N. Nikolov. LNCS, vol. 3843 (2005), pp. 117–128
51. M.B. Dillencourt, Realizability of Delaunay triangulations. Inform. Process. Lett. **33**(6), 283–287 (1990)
52. M.B. Dillencourt, W.D. Smith, Graph-theoretical conditions for inscribability and Delaunay realizability. Discr. Math. **161**(1–3), 63–77 (1996)
53. D. Dolev, H.W. Trickey, On linear area embedding of planar graphs. Technical report, Stanford University, Stanford, CA, 1981
54. P. Eades, Q. Feng, X. Lin, H. Nagamochi, Straight-line drawing algorithms for hierarchical graphs and clustered graphs. Algorithmica **44**(1), 1–32 (2006)
55. P. Eades, Q. Feng, H. Nagamochi, Drawing clustered graphs on an orthogonal grid. J. Graph Algorithm. Appl. **3**(4), 3–29 (1999)
56. P. Eades, S. Whitesides, The logic engine and the realization problem for nearest neighbor graphs. Theor. Comput. Sci. **169**(1), 23–37 (1996)
57. P. Eades, S. Whitesides, The realization problem for Euclidean minimum spanning trees in NP-hard. Algorithmica **16**(1), 60–82 (1996)
58. D. Eppstein, M.T. Goodrich, Succinct greedy graph drawing in the hyperbolic plane, in *Graph Drawing (GD'08)*, ed. by I.G. Tollis, M. Patrignani. LNCS, vol. 5417 (2009), pp. 14–25
59. I. Fáry, On straight line representation of planar graphs. Acta Sci. Math. **11**, 229–233 (1948)
60. S. Felsner, G. Liotta, S.K. Wismath, Straight-line drawings on restricted integer grids in two and three dimensions. J. Graph Algorithm. Appl. **7**(4), 363–398 (2003)
61. Q. Feng, Algorithms for drawing clustered graphs. Ph.D. thesis, The University of Newcastle, Australia, 1997
62. Q. Feng, R.F. Cohen, P. Eades, How to draw a planar clustered graph, in *Computing and Combinatorics (COCOON'95)*, ed. by D. Du, M. Li. LNCS, Xi'an, China, vol. 959 (1995), pp. 21–30
63. Q. Feng, R.F. Cohen, P. Eades, Planarity for clustered graphs, in *European Symposium on Algorithms (ESA'95)*, ed. by P.G. Spirakis. LNCS, Corfu, Greece, vol. 979 (1995), pp. 213–226
64. F. Frati, Straight-line drawings of outerplanar graphs in $O(dn\log n)$ area, in *Canadian Conference on Computational Geometry (CCCG'07)*, Carleton University, Ottawa, Canada 2007, pp. 225–228
65. F. Frati, On minimum area planar upward drawings of directed trees and other families of directed acyclic graphs. Int. J. Comput. Geom. Appl. **18**(3), 251–271 (2008)
66. F. Frati, Straight-line orthogonal drawings of binary and ternary trees, in *Graph Drawing (GD'07)*, ed. by S.-H. Hong, T. Nishizeki, W. Quan. LNCS, Sydney, Australia, vol. 4875 (2008), pp. 76–87
67. F. Frati, Lower bounds on the area requirements of series-parallel graphs. Discr. Math. Theor. Comput. Sci. **12**(5), 139–174 (2010)

68. F. Frati, M. Kaufmann, Polynomial area bounds for MST embeddings of trees. Tech. report RT-DIA-122-2008, Dept. of Computer Science and Automation, Roma Tre University, Rome, 2008
69. F. Frati, M. Patrignani, A note on minimum area straight-line drawings of planar graphs, in *Graph Drawing (GD'07)*, ed. by S.H. Hong, T. Nishizeki. LNCS, Sydney, Australia, vol. 4875 (2007), pp. 339–344
70. A. Garg, M.T. Goodrich, R. Tamassia, Planar upward tree drawings with optimal area. Int. J. Comput. Geom. Appl. **6**(3), 333–356 (1996)
71. A. Garg, A. Rusu, Area-efficient order-preserving planar straight-line drawings of ordered trees. Int. J. Comput. Geom. Appl. **13**(6), 487–505 (2003)
72. A. Garg, A. Rusu, Straight-line drawings of general trees with linear area and arbitrary aspect ratio, in *Computational Science and Its Applications (ICCSA'03)*, ed. by V. Kumar, M.L. Gavrilova, C.J.K. Tan, P. L'Ecuyer. LNCS, Montreal, Canada, vol. 2669 (2003), pp. 876–885
73. A. Garg, A. Rusu, Straight-line drawings of binary trees with linear area and arbitrary aspect ratio. J. Graph Algorithm. Appl. Rome, Italy, **8**(2), 135–160 (2004)
74. A. Garg, A. Rusu, Area-efficient planar straight-line drawings of outerplanar graphs. Discr. Appl. Math. **155**(9), 1116–1140 (2007)
75. A. Garg, R. Tamassia, Efficient computation of planar straight-line upward drawings, in *Graph Drawing (Proc. ALCOM Workshop on Graph Drawing)*, Sevres, Paris, 1994, pp. 298–306
76. A. Garg, R. Tamassia, On the computational complexity of upward and rectilinear planarity testing. SIAM J. Comput. **31**(2), 601–625 (2001)
77. M.T. Goodrich, G.S. Lueker, J.Z. Sun, C-planarity of extrovert clustered graphs, in *Graph Drawing (GD'05)*, ed. by P. Healy, N. Nikolov. LNCS, Limerick, Ireland, vol. 3843 (2005), pp. 211–222
78. M.T. Goodrich, D. Strash, Succinct greedy geometric routing in the Euclidean plane, in *Algorithms and Computation (ISAAC'09)*, ed. by Y. Dong, D.-Z. Du, O.H. Ibarra. LNCS, Hawai, USA, vol. 5878 (2009), pp. 781–791
79. M.T. Goodrich, C.G. Wagner, A framework for drawing planar graphs with curves and polylines. J. Algorithms **37**(2), 399–421 (2000)
80. C. Gutwenger, M. Jünger, S. Leipert, P. Mutzel, M. Percan, R. Weiskircher, Advances in c-planarity testing of clustered graphs, in *Graph Drawing (GD'02)*, ed. by S.G. Kobourov, M.T. Goodrich. LNCS, Irvine, California, USA, vol. 2528 (2002), pp. 220–235
81. C. Gutwenger, P. Mutzel, Planar polyline drawings with good angular resolution, in *Graph Drawing (GD'98)*, ed. by S. Whitesides. LNCS, Montréal, Canada, vol. 1547 (1998), pp. 167–182
82. X. He, Grid embedding of 4-connected plane graphs. Discr. Comput. Geom. **17**(3), 339–358 (1997)
83. X. He, H. Zhang, Succinct convex greedy drawing of 3-connected plane graphs, in *Symposium on Discrete Algorithms (SODA'11)*, San Francisco, California, USA, 2011, pp. 1477–1486
84. M.D. Hutton, A. Lubiw, Upward planarity testing of single-source acyclic digraphs. SIAM J. Comput. **25**(2), 291–311 (1996)
85. V. Jelínek, E. Jelínková, J. Kratochvíl, B. Lidický, Clustered planarity: embedded clustered graphs with two-component clusters, in *Graph Drawing (GD'08)*, ed. by I.G. Tollis, M. Patrignani. LNCS, Heraklion, Crete, Greece, vol. 5417 (2008), pp. 121–132
86. V. Jelínek, O. Suchý, M. Tesař, T. Vyskočil, Clustered planarity: clusters with few outgoing edges, in *Graph Drawing (GD'08)*, ed. by I.G. Tollis, M. Patrignani. LNCS, Heraklion, Crete, Greece, vol. 5417 (2008), pp. 102–113
87. E. Jelínková, J. Kára, J. Kratochvíl, M. Pergel, O. Suchý, T. Vyskočil, Clustered planarity: small clusters in cycles and Eulerian graphs. J. Graph Algorithm Appl. **13**(3), 379–422 (2009)
88. G. Kant, Drawing planar graphs using the canonical ordering. Algorithmica **16**(1), 4–32 (1996)
89. G. Kant, X. He, Regular edge labeling of 4-connected plane graphs and its applications in graph drawing problems. Theor. Comput. Sci. **172**(1–2), 175–193 (1997)

90. M. Kaufmann, Polynomial area bounds for MST embeddings of trees, in *Graph Drawing (GD'07)*, ed. by S.H. Hong, T. Nishizeki, W. Quan. LNCS, Sydney, Australia, vol. 4875 (2007), pp. 88–100
91. M. Kaufmann, D. Wagner (eds.), *Drawing Graphs, Methods and Models*. Lecture Notes in Computer Science (Springer-Verlag, New York/Heidelberg, 2001)
92. S.K. Kim, Simple algorithms for orthogonal upward drawings of binary and ternary trees, in *Canadian Conference on Computational Geometry (CCCG'95)*, Quebec City, Quebec, Canada, 1995, pp. 115–120
93. R. Kleinberg, Geographic routing using hyperbolic space, in *IEEE Conference on Computer Communications (INFOCOM'07)*, Anchorage, Alaska, USA, 2007, pp. 1902–1909
94. T. Leighton, A. Moitra, Some results on greedy embeddings in metric spaces. Discr. Comput. Geom. **44**(3), 686–705 (2010)
95. W. Lenhart, G. Liotta, Proximity drawings of outerplanar graphs, in *Graph Drawing (GD'96)*, ed. by S.C. North. LNCS, Berkeley, California, USA, vol. 1190 (1996), pp. 286–302
96. G. Liotta, Computing proximity drawings of graphs. Ph.D. thesis, University of Rome "La Sapienza." Rome, 1995
97. G. Liotta, R. Tamassia, I.G. Tollis, P. Vocca, Area requirement of Gabriel drawings, in *Algorithms and Complexity (CIAC'97)*, ed. by G.C. Bongiovanni, D.P. Bovet, G. Di Battista. LNCS, Rome, Italy, vol. 1203 (1997), pp. 135–146
98. A. Lubiw, N. Sleumer, Maximal outerplanar graphs are relative neighborhood graphs, in *Canadian Conference on Computational Geometry (CCCG'93)*, Waterloo, Ontario, Canada, 1993, pp. 198–203
99. K. Miura, S.-I. Nakano, T. Nishizeki, Grid drawings of 4-connected plane graphs. Discr. Comput. Geom. **26**(1), 73–87 (2001)
100. D. Mondal, R.I. Nishat, Md. S. Rahman, Md. J. Alam, Minimum-area drawings of plane 3-trees, in *Canadian Conference on Computational Geometry (CCCG'10)*, Manitoba, Canada, 2010, pp. 191–194
101. C.L. Monma, S. Suri, Transitions in geometric minimum spanning trees. Discr. Comput. Geom. **8**, 265–293 (1992)
102. T. Nishizeki, N. Chiba, *Planar Graphs: Theory and Algorithms* (North Holland, Amsterdam, 1988)
103. C.H. Papadimitriou, D. Ratajczak, On a conjecture related to geometric routing. Theor. Comput. Sci. **344**(1), 3–14 (2005)
104. A. Papakostas, Upward planarity testing of outerplanar dags, in *Graph Drawing*, ed. by R. Tamassia, I.G. Tollis. LNCS, Springer, vol. 894 (1994), pp. 298–306
105. P. Penna, P. Vocca, Proximity drawings in polynomial area and volume. Comput. Geom.: Theor. Appl. **29**(2), 91–116 (2004)
106. S. Rabinowitz, $O(n^3)$ bounds for the area of a convex lattice n-gon. Geombinatorics **2**, 85–88 (1993)
107. A. Rao, C.H. Papadimitriou, S. Shenker, I. Stoica, Geographic routing without location information, in *ACM International Conference on Mobile Computing and Networking (MOBICOM'03)*, ed. by D.B. Johnson, A.D. Joseph, N.H. Vaidya, San Diego, California, USA, 2003, pp. 96–108
108. R.C. Read, A new method for drawing a graph given the cyclic order of the edges at each vertex. Congressus Numerantium **56**, 31–44 (1987)
109. G. Rote, Strictly convex drawings of planar graphs, in *Symposium on Discrete Algorithms (SODA'05)*, Vancouver, BC, Canada, 2005, pp. 728–734
110. W. Schnyder, Embedding planar graphs on the grid, in *Symposium on Discrete Algorithms (SODA'90)*, San Francisco, California, USA, 1990, pp. 138–148
111. W. Schnyder, W. Trotter, Convex drawings of planar graphs. Abstr. Am. Math. Soc. 92T-05-135 (1992)
112. C.S. Shin, S.K. Kim, K.Y. Chwa, Area-efficient algorithms for straight-line tree drawings. Comput. Geom.: Theor. Appl. **15**(4), 175–202 (2000)
113. S.K. Stein, Convex maps. Am. Math. Soc. **2**, 464–466 (1951)

114. M.A. Stern, Ueber eine zahlentheoretische funktion. J. Reine Angew. Math. **55**, 193–220 (1858)
115. R. Tamassia (ed.), *Handbook of Graph Drawing and Visualization* (CRC Press), to appear
116. R. Tamassia, I.G. Tollis, A unified approach to visibility representation of planar graphs. Discr. Comput. Geom. **1**, 321–341 (1986)
117. C. Thomassen, Planarity and duality of finite and infinite graphs. J. Comb. Theor. Ser. B **29**(2), 244–271 (1980)
118. C. Thomassen, Plane representations of graphs, in *Progress in Graph Theory* (Academic, 1984), pp. 43–69
119. W.T. Tutte, Convex representations of graphs. Lond. Math. Soc. **10**, 304–320 (1960)
120. W.T. Tutte, How to draw a graph. Lond. Math. Soc. **13**(3), 743–768 (1963)
121. L.G. Valiant, Universality considerations in VLSI circuits. IEEE Trans. Comput. **30**(2), 135–140 (1981)
122. K. Wagner, Bemerkungen zum vierfarbenproblem. Jahresbericht. German. Math.-Verein **2**, 26–32 (1936)
123. D. Woods, Drawing planar graphs. Ph.D. thesis, Stanford University, Stanford, CA, 1982
124. H. Zhang, Planar polyline drawings via graph transformations. Algorithmica **57**(2), 381–397 (2010)
125. H. Zhang, S. Sadasivam, On planar polyline drawings, in *Graph Drawing*, ed. by S.-H. Hong, T. Nishizeki, W. Quan. LNCS, Springer, vol. 4875 (2008), pp. 213–218
126. X. Zhou, T. Hikino, T. Nishizeki, Small grid drawings of planar graphs with balanced bipartition, in *Algorithms and Computation (WALCOM'10)*, ed. by Md. S. Rahman, S. Fujita. LNCS, Dhaka, Bangladesh, vol. 5942 (2010), pp. 47–57

The Crossing-Angle Resolution in Graph Drawing

Walter Didimo and Giuseppe Liotta

Abstract The crossing-angle resolution of a drawing of a graph measures the smallest angle formed by any pair of crossing edges. In this chapter, we survey some of the most recent results and discuss the current research agenda on drawings of graphs with good crossing-angle resolution.

1 Introduction

An emerging line of research in graph drawing studies nonplanar drawings of graphs. The growing interest into this subject can be partly justified by the following two observations.

The planarity handicap. A large part of the existing literature on graph drawing showcases elegant algorithms and sophisticated data structures under the assumption that the input graph is planar (see, e.g., [13, 36, 37]). This assumption finds its natural justification in the numerous experimental studies establishing that the human ability of understanding a diagram is strongly affected by the number of edge crossings (see, e.g., [44, 45, 50]). Unfortunately, many graphs are nonplanar in practice. Hence, there is an increasing need for efficient and effective systems that can visually explore the torrents of nonplanar relational data sets generated in a variety of application domains (see, e.g., [48]).

Turán-type problems and algorithmic problems. The study of Turán-type problems about the edge density of graphs with forbidden subgraphs has a long tradition in combinatorial and discrete geometry (see, e.g. [40]). The need for designing new drawing paradigms for nonplanar representations of graphs leads to "Turán-type

W. Didimo (✉) • G. Liotta
Dipartimento di Ingegneria Elettronica e dell'Informazione, Università degli Studi di Perugia, Perugia, Italy
e-mail: didimo@diei.unipg.it; liotta@diei.unipg.it

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_10,

algorithmic problems," that is, to the problem of designing algorithms that compute nonplanar drawings where subdrawings of certain types are forbidden because they are considered visually confusing.

In this context, a recent series of experiments by Huang et al. [32, 33, 35] shows that crossing edges significantly affect the human understanding of a diagram if they form acute angles, while those edge crossings that form angles from about $\frac{\pi}{3}$ to $\frac{\pi}{2}$ guarantee good readability properties. As a result, a growing number of papers in the graph drawing literature study nonplanar representations where such "sharp angle crossings" are forbidden. The aim of this chapter is to survey this literature, emphasizing both combinatorial and algorithmic results.

The terminology throughout the chapter is as follows. A *drawing* of a graph G (a) injectively maps each vertex u of G to a point p_u in the plane, (b) maps each edge (u,v) of G to a Jordan arc connecting p_u and p_v that does not pass through any other vertex, (c) is such that any two edges have at most one point in common. The *angle resolution* of a drawing D of a graph is the minimum angle formed by any pair of crossing edges in G. By using the above terminology, this chapter considers the following types of drawings of graphs.

- Drawings with *optimum* crossing-angle resolution, i.e., where every pair of crossing edges forms angle $\frac{\pi}{2}$. We call these drawings *right angle-crossing drawings* (or simply *RAC drawings*).
- Drawings that have crossing-angle resolution (at least) a given α, for $0 < \alpha < \pi/2$.

The remainder of the chapter is organized as follows. Section 2 considers extremal problems on straight-line RAC drawings; straight-line drawings of graphs are often called *geometric graphs*. Hence, straight-line RAC drawings will be also referred to as *geometric RAC graphs*. Section 3 considers the problem of recognizing and computing geometric RAC graphs. Variants and generalizations of geometric RAC graphs are considered in Sect. 4; we shall talk about drawings where bends along the edges are allowed and about drawings such that, for a given angle α where $0 < \alpha < \pi/2$, the edge crossings form angles that are either *exactly* α or *at least* α. Finally, Sect. 5 discusses future research directions.

2 Geometric RAC Graphs

The notion of RAC drawings was first introduced in [19, 21], where the following Turán-type result is also proved. The first part of the theorem provides an upper bound on the number of edges in a geometric RAC graph, while the second part implies that this bound is tight.

Theorem 1 ([19, 21]). *Every geometric RAC graph with n vertices has at most $4n-10$ edges. Also, for any $k \geq 3$ there exists a RAC graph with $n = 3k-5$ vertices and $4n-10$ edges.*

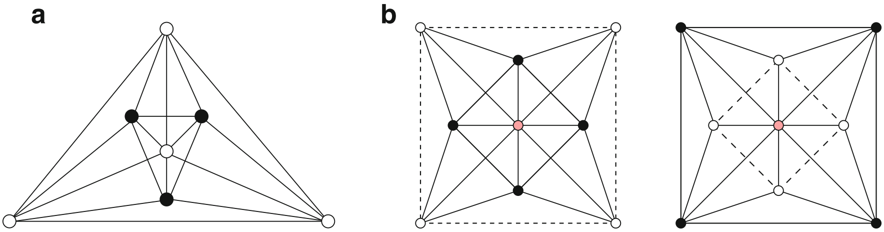


Fig. 1 (**a**) A geometric RAC graph with $n = 7$ vertices and $m = 4n - 10$ edges. (**b**) Two different straight-line RAC drawings of the square antiprism graph; the two drawings have a different combinatorial embedding

The proof of the upper bound in Theorem 1 is based on the following ideas (see [19, 21] for details). Let D be a geometric RAC graph; first, it is shown that the edges of D can be colored with three colors, say red, blue, and green, in such a way that red edges do not cross any other edge in D, while each blue (green) edge always crosses some green (blue) edges. Such a coloring defines two geometric graphs, D_{rb} and D_{rg}, the first consisting of the red and blue edges, and the second consisting of the red and green edges. Both D_{rb} and D_{rg} are planar and share the external face. This immediately implies that D has at most $6n - 12$ edges. The upper bound $4n - 10$ is then proved with a case analysis on the degree of the external face of D_{rb} and of its adjacent internal faces. More precisely, it is shown that in a geometric RAC graph of maximum edge density, every internal face of D_{rb} has at least two red edges, while the external face consists of red edges only. This observation, together with counting arguments based on Euler's formula, leads to proving that the number of edges of a geometric RAC graph with n vertices cannot exceed $4n - 10$. In order to show the tightness of this upper bound, it is proved that for any $k \geq 3$, there exists a graph G_k with $n = 3k - 5$ vertices and $4n - 10$ edges that admits a straight-line RAC drawing. Graph G_k is constructed as follows (refer also to Fig. 1a): Start from an embedded maximal planar graph with k vertices and add to this graph its dual planar graph without the face node corresponding to the external face (in Fig. 1a the primal graph has white vertices and the dual graph has black vertices). Also, for each face node u, add to G_k three edges that connect u to the three vertices of the face associated with u. The fact that G_k is realizable as a geometric RAC graph is a consequence of a disk-packing theorem of Brightwell and Scheinerman [11].

A *maximally dense* RAC graph is a geometric RAC graph having n vertices and $4n - 10$ edges. In [27], the relationship between maximally dense RAC graphs and another well-known family of geometric graphs is studied. A graph is *1-planar* if it admits a drawing where every edge is crossed at most once. A list of references about 1-planar graphs is included in [10, 29, 38, 42, 47].

Theorem 2 ([27]). *Every maximally dense RAC graph is* 1*-planar. Also, for every integer i such that* $i \geq 0$*, there exists a* 1*-planar graph with* $n = 8 + 4i$ *vertices and* $4n - 10$ *edges that is not a RAC graph.*

Dujmović et al. [24, 25] present an alternative proof of the upper bound given in Theorem 1. Instead of exploiting edge-coloring arguments, their proof relies on charging techniques similar to those used by Ackerman and Tardos [3] and Ackerman [1], but it still applies a case analysis on the degree of the external face and of its adjacent faces, as in the proof of Theorem 1. Roughly speaking, charging techniques are based on assigning suitable *charges* to the faces and/or to the vertices of a plane subdivision defined by a nonplanar drawing. These charges can be expressed, for example, as functions of the degree of faces and vertices of the subdivision. Charges are often redistributed to guarantee desired properties and so that the sum of all charges remains the same. Using Euler's formula, it is possible to establish equations that relate the sum of all charges to some functions of the number of vertices of the graph, and the number of edges to the number of vertices of the graph.

In addition to (and partially motivated by) Theorem 1, many recent papers concerning drawings of graphs with good crossing-angle resolution have been published. Essentially, the following two questions have been investigated.

- What is the complexity of recognizing geometric RAC graphs?
- Can one draw nonsparse graphs with good crossing-angle resolution either by allowing crossing angles smaller than $\frac{\pi}{2}$ or by allowing bends along the edges, or both?

The next sections both survey the most relevant known results about the two problems above and discuss some future research directions.

3 Complexity Questions About Geometric RAC Graphs

Argyriou et al. [7] show that recognizing those graphs that are realizable as geometric RAC graphs is computationally difficult.

Theorem 3 ([7]). *Let G be a graph. It is $\mathcal{NP}$-hard to decide whether G admits a straight-line RAC drawing.*

The proof of Theorem 3 uses a reduction from the well-known 3-SAT problem. The building block of the reduction is a small graph of nine vertices, called an *augmented square antiprism graph*, whose straight-line RAC drawings have only two possible combinatorial embeddings (see Fig. 1b).

One can wonder whether results similar to Theorem 3 hold if additional geometric constraints are added to a geometric RAC graph. Angelini et al. [4, 5] study straight-line *upward RAC drawings* of digraphs, that is, RAC drawings where the edges are monotone in the upward direction according to their orientation. They show that the problem of recognizing those digraphs that admit a straight-line upward RAC drawing remains computationally difficult.

Theorem 4 ([4]). *Let G be a digraph. It is $\mathcal{NP}$-hard to decide whether G admits a straight-line upward RAC drawing.*

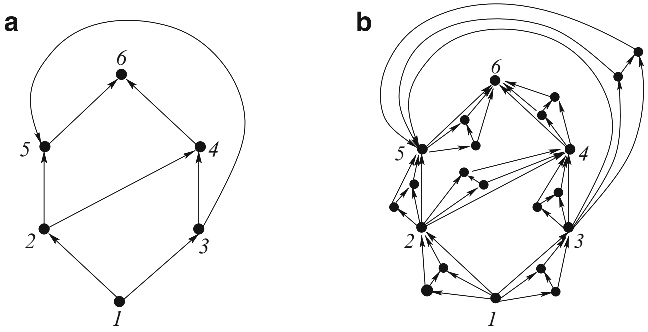


Fig. 2 (**a**) A planar digraph G. (**b**) A digraph G' that has a straight-line upward RAC drawing if and only if G is upward planar. In this case, G is not upward planar and G' does not have a straight-line upward RAC drawing

The proof of Theorem 4 is based on the following idea. Consider any planar digraph G and construct a digraph G' obtained from G by replacing each edge (u,v) with the complete graph K_4 whose edges are oriented from u to v (see Fig. 2). It is proved that G is upward planar if and only if G' admits a straight-line upward RAC drawing [4]. Since the problem of deciding whether a digraph is upward planar is $\mathcal{NP}$-hard [31], the above reduction immediately implies the hardness of deciding whether a digraph has a straight-line upward RAC drawing.

On the positive side, Di Giacomo et al. [16] prove that recognizing the bipartite graphs that admit a straight-line RAC drawing where vertices of different bipartite sets lie on two distinct horizontal levels can be recognized in polynomial time.

Theorem 5 ([16]). *Let $G = (V_1, V_2, E)$ be a bipartite graph. There exists a linear-time algorithm that tests whether G admits a straight-line RAC drawing such that the vertices of V_1 and those of V_2 lie on two distinct horizontal layers. Also, such a drawing can be computed in linear time if it exists.*

Theorem 5 exploits a characterization of the family of geometric RAC graphs on two layers. Roughly speaking, the graphs in this family consist of nontrivial biconnected components that are spanning subgraphs of *ladders* plus tree-like components with special properties. A ladder is a biconnected bipartite graph consisting of two paths of the same length $\langle u_1, u_2, \ldots, u_{\frac{n}{2}} \rangle$ and $\langle v_1, v_2, \ldots, v_{\frac{n}{2}} \rangle$, plus the edges (u_i, v_i) $(i = 1, 2, \ldots, \frac{n}{2})$ (see Fig. 3). We observe that since ladders are the densest graphs admitting a straight-line RAC drawing on two layers, the maximum number of edges of a geometric RAC graph on two layers is $1.5n - 2$ (i.e., the number of vertices of a ladder).

In the same paper, the following $\mathcal{NP}$-hardness result is also proved.

Theorem 6 ([16]). *Let G be a bipartite graph and let k be a positive integer. It is $\mathcal{NP}$-complete to decide whether G admits a subgraph with at least k edges that admits a straight-line RAC drawing on two layers.*

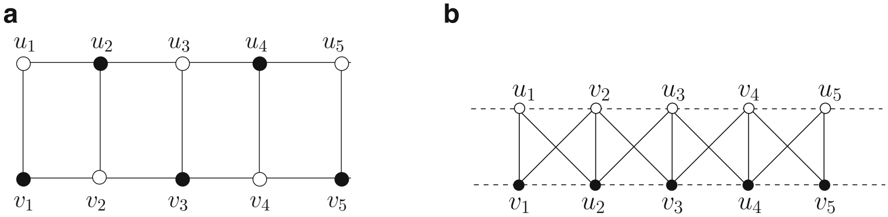


Fig. 3 (**a**) A ladder. (**b**) A straight-line RAC drawing of the ladder on two layers

The complete bipartite graphs that admit a straight-line RAC drawing with no constraint on the vertex position are characterized in [20]. The characterization immediately leads to an efficient recognition algorithm for the straight-line RAC drawable bipartite graphs.

Theorem 7 ([20]). *A complete bipartite graph K_{n_1,n_2} $(n_1 \leq n_2)$ admits a straight-line RAC drawing if either $n_1 \leq 2$, or $n_1 = 3$ and $n_2 \leq 4$.*

We conclude this section by observing that complexity issues in graph drawing are often not just evaluated in computational terms but also in terms of *visual complexity*. The visual complexity of a drawing can be expressed by means of different parameters, often called *aesthetics*, such as, for example, the area requirement, the total edge length, and the angular resolution. Van Kreveld [49] has recently studied how much better a straight-line RAC drawing of a planar graph can be than any planar drawing of the same graph. He defines the following measure, called the *quality ratio*:

$$QR(\Phi) = \sup_{G \text{ planar}} \frac{\Phi_{\text{RAC}}(G)}{\Phi_{\text{planar}}(G)},$$

where $\Phi_{\text{RAC}}(G)$ is the optimum value of measure Φ over all RAC drawings of G, and $\Phi_{\text{planar}}(G)$ is the optimum value of measure Φ over all planar drawings of G. The optimum value of a measure can be either its maximum or its minimum value, depending on the kind of measure. In particular, denoted by Φ^A, Φ^E, and Φ^α the area requirement, the total edge length, and the angular resolution, respectively, the following result has been proven.

Theorem 8 ([49]). $QR(\Phi^A) = QR(\Phi^E) = \infty$. $QR(\Phi^\alpha) \geq 2.75$.

The optimum values for Φ^A and Φ^E are their minimum values, while the optimum value for Φ^α is its maximum value.

Theorem 8 essentially implies that in many cases a planar graph G can be drawn nonplanarly, with optimum crossing-angle resolution, in order to improve some other important aesthetics rather than the number of crossings. On the negative side, however, there are results showing that RAC drawings do not always help in this sense. Angelini et al. [5] prove that the area requirement of geometric RAC graphs is $\Omega(n^2)$; that is, there exist infinitely many graphs whose straight-line RAC

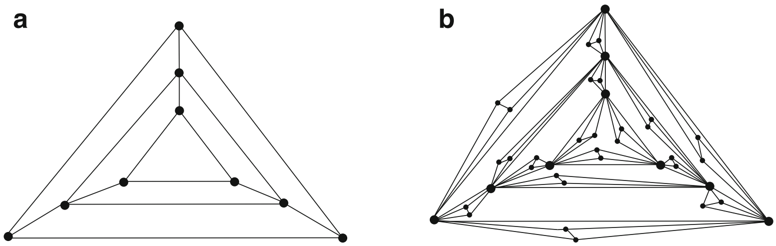


Fig. 4 (**a**) A nested-triangles graph G. (**b**) The graph G' obtained by replacing each edge (a,b) of G with a copy of K_4

drawings always require quadratic area. More precisely, consider a nested-triangles graph G, that is, a triconnected graph composed of $\frac{n}{3}$ 3-cycles nested one into the other (see Fig. 4a). Graph G is known to require $\Omega(n^2)$ area in any straight-line planar drawing [12]. Replace each edge (a,b) of G with a copy of K_4 , by identifying vertices a and b of G with vertices u and v of K_4 , respectively. Let G' be the resulting planar graph (see Fig. 4b). G' has $O(n)$ vertices since, for every edge of G, two new vertices are introduced in G. The following theorem is proved.

Theorem 9 ([5]). *Any straight-line RAC drawing of G' requires $\Omega(n^2)$ area.*

The idea of the proof is based on the observation that G is a subgraph of G' and that any straight-line RAC drawing D' of G' is such that no two edges of G in D' cross each other. This implies that if D' had subquadratic area, then G would admit a straight-line planar drawing with subquadratic area, which is impossible.

4 Curve Complexity and Crossing-Resolution Tradeoffs

Theorem 1 restricts the geometric RAC graphs to being sparse. In order to compute drawings of graphs with good crossing-angle resolution and with superlinear edge density, variants and relaxations of RAC graphs must be considered. For example, one can study either RAC drawings where bends along the edges are allowed, or drawings where the edge crossings may not be orthogonal. The next subsection concentrates on RAC drawings with bends along the edges. Section 4.2 considers drawings where crossing angles can be smaller than $\frac{\pi}{2}$.

4.1 Allowing Bends Along the Edges

The *curve complexity* of a drawing D is the maximum number of bends along an edge of D (therefore, straight-line drawings have curve complexity 0). In [19, 21] the study of RAC drawings with bent edges is initiated. The following result is given.

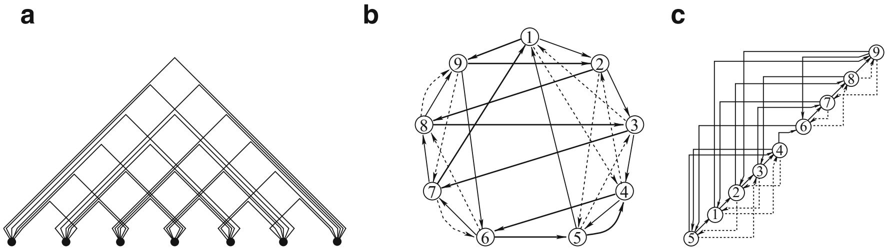


Fig. 5 (**a**) A RAC drawing of K_7 with curve complexity 3. (**b**) A regular directed multigraph G' with in-degree and out-degree equal to 3 and its cycle covers C_1, C_2, and C_3. The edges of C_1 are represented by *solid thin lines*, the ones of C_2 are represented by *solid thick lines*, and the ones of C_3 are represented by *dashed lines*. (**c**) A RAC drawing of G' with two bends per edge constructed by the algorithm of Theorem 13

Theorem 10 ([19, 21]). *Every graph admits a RAC drawing with curve complexity 3.*

There are different ways of proving Theorem 10. For example, one can place all vertices on a horizontal line with an arbitrary order, and use exactly three bends per edge so that any two crossing segments have slopes 1 and -1, respectively (see Fig. 5a). In [19, 21] it is also proved that not all graphs can be drawn with optimum crossing-angle resolution and curve complexity smaller than 3; indeed, it is shown that RAC drawings with curve complexity 1 have $O(n^{4/3})$ edges, while those with curve complexity 2 have $O(n^{7/4})$ edges. These last upper bounds on the edge density are significantly improved by Arikushi et al. [8].

Theorem 11 ([8]). *An RAC drawing with n vertices and curve complexity 1 has at most* $6.5n - 13$ *edges.*

Theorem 12 ([8]). *An RAC drawing with n vertices and curve complexity 2 has less than* $74.2n$ *edges.*

The proof of Theorem 11 still relies on charging arguments on the plane subdivision defined by an RAC drawing with curve complexity 1. The proof of Theorem 12 is based on a completely different argument. It combines the new concept of *block graph*, defined on the set of crossing segments in an RAC drawing with curve complexity 2, with the strongest-known version of the *crossing lemma*, due to Pach et al. [41].

Lemma 4.1 ([41]). *Let G be a graph with n vertices and m edges. If* $m \geq \frac{103}{6}n \approx 17.167n$*, then*

$$cr(G) \geq \frac{1024}{31827} \cdot \frac{m^3}{n^2} \approx 0.032\frac{m^3}{n^2}.$$

In the lemma above, $cr(G)$ denotes the minimum number of crossings in a plane drawing of G.

Angelini et al. [4, 5] study RAC drawings with bends along the edges for graphs with a bounded vertex degree. They prove the following results.

Theorem 13 ([4]). *Let G be a graph with n vertices and vertex degree at most 6. There always exists a RAC drawing of G with curve complexity 2 and area $O(n^2)$. Such a drawing can be computed in $O(n)$ time.*

Theorem 14 ([4]). *Let G be a graph with n vertices and vertex degree at most three. There always exists an RAC drawing of G with curve complexity 1 and area $O(n^2)$. Such a drawing can be computed in $O(n)$ time.*

The algorithms in Theorems 13 and 14 are based on the decomposition of a regular directed multigraph into directed 2-factors. A *2-factor* of an undirected graph G is a spanning subgraph of G consisting of vertex-disjoint cycles (see, e.g., [9]). Analogously, a *directed 2-factor* of a directed graph is a spanning subgraph consisting of vertex-disjoint directed cycles. The decomposition of a regular directed multigraph into directed 2-factors follows from a classical result for undirected graphs, stating that "a regular multigraph of degree $2k$ has k edge-disjoint 2-factors" [43]. Given a graph G with vertex degree at most Δ, Eades, Symvonis, and Whitesides [28] show how to construct a digraph G' such that (a) each vertex of G has in-degree and out-degree at $d = \lceil \Delta/2 \rceil$, (b) G is a subgraph of the undirected underlying graph of G', (c) the edges of G' can be partitioned into d edge-disjoint directed 2-factors. If $\Delta = 6$, the algorithm in Theorem 13 constructs an RAC drawing of G' with curve complexity 2, placing all vertices on a line ℓ of slope 1, almost all the edges of one of the three 2-factors along ℓ, and one of the edge sets of the two remaining 2-factors is "above" ℓ while the other is "below" ℓ (see, Fig. 5b, c). A similar technique is used if $\Delta = 3$.

Area-curve complexity tradeoffs of RAC drawings are further studied by Di Giacomo et al. [17, 18]. They observe that the constructive technique described for proving Theorem 10 gives rise to RAC drawings with area $O(n^4)$, and prove that by allowing curve complexity larger than 3, one can reduce the area of an RAC drawing.

Theorem 15 ([17,18]). *Every graph admits an RAC drawing with curve complexity 4 and area $O(n^3)$.*

The proof of Theorem 15 is constructive. It describes how to compute an RAC drawing of the complete graph K_n with exactly four bends per edge and cubic area. Arbitrarily number the vertices of K_n from 0 to $n-1$. Vertex i with $0 \le i \le n-1$ is placed at point $p_i = (in-3, 2n)$. For each pair of vertices i and j, with $i < j$, the four bends of edge (i, j) are placed at the following points (in this order):

- $a_{i,j} = (in-2, (j-i)-1+2(n-1))$,
- $b_{i,j} = (in, (j-i)-1)$,
- $c_{i,j} = (jn-(j-i), 2(j-i)-1)$,
- $d_{i,j} = (jn-(j-i)-2, 2(j-i)-1+2(n-1))$.

An RAC drawing of K_6 constructed with this technique is shown in Fig. 6.

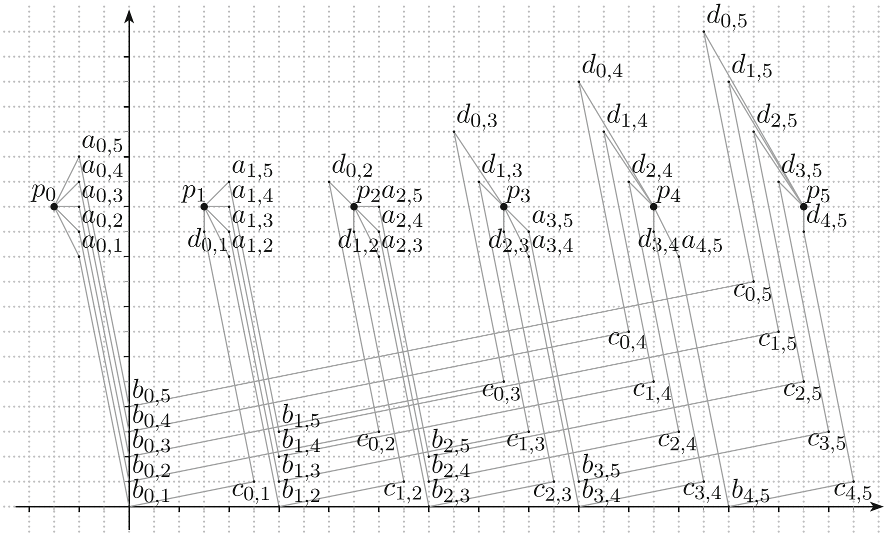


Fig. 6 An *RAC* drawing of K_6 with curve complexity 4

4.2 Drawings with Nonorthogonal Edge Crossings

In the last couple of years, several papers have been devoted to drawings of graphs where the crossing angles are not required to be orthogonal, but a lower bound on the crossing-angle resolution is given as part of the input and the curve complexity is kept low. Since the existing papers often adopt a nonuniform and somewhat confusing notation, for the sake of clarity, we redefine these drawings as follows.

Let α be an angle value such that $0 < \alpha < \frac{\pi}{2}$. An *ACE$\alpha$ drawing* is a polyline drawing such that any two edges cross at an angle that is *exactly* α. An *ACLα graph* is a polyline drawing such that any two edges cross at an angle that is *at least* α. When the edges of an ACEα drawing (ACLα drawing) are all straight-line segments, the drawing is a geometric graph that we also call a *geometric ACEα graph* (a *geometric ACLα graph*). For completeness, we recall that ACLα drawings have been independently introduced by Di Giacomo et al. [17, 18], who call them *LAC$_\alpha$ drawings*, and by Dujmović et al. [24, 25], who call them *αAC drawings*. Ackerman et al. [2] introduced and studied ACEα drawings with the name of $\alpha AC_b^=$ (b is used to denote the maximum number of bends per edge in the drawing). Dujmović et al. [24, 25] prove the following result.

Theorem 16 ([24, 25]). *A geometric ACLα graph with n vertices has at most $\frac{\pi}{\alpha}(3n-6)$ edges.*

The idea behind the proof of Theorem 16 uses a bucketing argument. Namely, assume for simplicity that $k = \frac{\pi}{\alpha}$ is an integer number, and let D be a geometric ACLα graph. Partition the edge set of D into k subsets called *buckets*; bucket i

$(1 \le i \le k)$ contains all edges that have their *direction* in the interval $[\alpha(i-1), \alpha i)$. The direction of an edge e is the angle formed by a horizontal line and a line that contains e in the counterclockwise direction; it ranges in the interval $[0, \pi)$. Denote by D_i the subgraph of D consisting of the edges in the bucket i only. Since the crossing-angle resolution of D is at least α, each D_i is planar; indeed, two crossing edges of D_i would form an angle smaller than α. It follows that each D_i has at most $3n-6$ edges, and hence the number of edges of D is at most $k(3n-6) = \frac{\pi}{\alpha}(3n-6)$. The proof is slightly refined with some probabilistic arguments if k is not an integer, but the bucketing argument remains the same.

Ackerman et al. [2] prove a similar result as the one of Theorem 16 for geometric ACEα graphs.

Theorem 17 ([2]). *A geometric ACEα graph with n vertices has at most* $3(3n-6)$ *edges.*

The proof of Theorem 17 is based on an edge-coloring argument that generalizes the one used for proving the $4n-10$ bound for geometric RAC graphs given in Theorem 1. Namely, suppose that D is a geometric ACEα graph with n vertices. One can partition the set of edges of D into maximal subsets S_i of pairwise parallel edges. Let $\mathcal{S} = \{S_i\}$ and let $G_{\mathcal{S}} = (\mathcal{S}, E_{\mathcal{S}})$ be a graph such that $(S_i, S_j) \in E_{\mathcal{S}}$ if and only if the direction of the edges in S_i differs from the direction of the edges in S_j by an angle α. Since each vertex of $G_{\mathcal{S}}$ has degree at most 2, the vertex set of $G_{\mathcal{S}}$ is 3-colorable. This vertex coloring corresponds to an edge coloring of D such that all the edges with the same color do not cross each other, and hence they induce a planar graph. Thus, the total number of edges of D is at most $3(3n-6)$. Concerning ACEα drawings with curve complexity larger than 0, Ackerman et al. prove the following results.

Theorem 18 ([2]). *An ACEα drawing with n vertices and curve complexity 1 has at most* $27n$ *edges.*

Theorem 19 ([2]). *An ACEα drawing with n vertices and curve complexity 2 has at most* $477n$ *edges.*

The key idea for the proofs of Theorems 18 and 19 is to partition the set of edge segments into two subsets: The *end-segments* are those incident to some vertex of the drawing; the *middle-segments* are the remaining edge segments. Then the bounds are derived counting the number of crossings that involve either two end segments or an end segment and a middle segment. We remark that a similar argument, combined with the crossing lemma, has been previously used in [19, 21] to derive the subquadratic bounds on the number of edges of RAC drawings with curve complexities 1 and 2.

Theorems 12 and 19 imply that curve complexity 2 is not sufficient to draw all graphs when the crossing angle is required to be a given $\alpha \in (0, \frac{\pi}{2}]$. On the other hand, three bends per edge always suffice, as the proof of Theorem 10 immediately extends to any other value of α (consider edges with middle segments that form an angle of $\pm\frac{\alpha}{2}$ with the horizontal line).

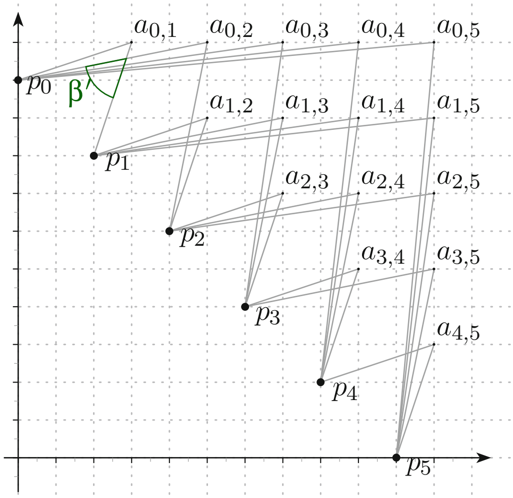


Fig. 7 An ACLα-drawing of K_6 with curve complexity 1 and $\alpha = 0.93$. In this drawing $c = 2$ (which guarantees that the crossing resolution is at least 0.93). Indeed, the smallest crossing angle in the drawing is β', and its value is 1.05

The edge density/curve complexity tradeoff turns out to be quite different for ACLα drawings. Namely, Di Giacomo et al. [17, 18] prove the following.

Theorem 20 ([17, 18]). *Every graph with n vertices has an ACLα drawing with curve complexity 1 and area $O(n^2)$.*

The proof of Theorem 20 describes how to construct an ACLα drawing of K_n with exactly one bend per edge and area $O(n^2)$. Let c be an integer such that the angle between a line of slope $c+1$ and a line of slope $\frac{1}{c+1}$ is larger than α. Arbitrarily number the vertices of K_n from 0 to $n-1$. Vertex i $(0 \leq i \leq n-1)$ is placed at point $p_i = (ic, (n-i-1)c)$. For each pair of vertices i and j, with $i < j$, the bend of edge (i, j) is placed at point $a_{i,j} = (jc+1, (n-i-1)c+1)$. Figure 7 shows an example of such a drawing for $n = 6$ and $c = 2$. Finally, the optimality of the area upper bound is an immediate consequence of Theorem 16, which implies that any ACLα drawing of a complete graph must have a quadratic number of edges with at least one bend (and therefore any drawing contains at least a quadratic number of points that represent such bends).

It may be worth comparing Theorem 15 with Theorem 20: While the best general upper bound on the area of an RAC drawing is $O(n^3)$ and requires curve complexity 4, for any arbitrarily small $\varepsilon > 0$ one can compute an ACLα drawing with $\alpha = \frac{\pi}{2} - \varepsilon$ having curve complexity 1 and quadratic area.

5 Future Research Directions

This chapter has surveyed recent results about nonplanar drawings with good crossing-angle resolution. The study of the subject is partly motivated by applications of information visualization and partly by its interest in the fields of

Table 1 Summary of Turán-type results for RAC, ACEα, and ACLα drawings

	Straight-line	1 bend/edge	2 bends/edge	3 bends/edge
RAC drawing	$M = 4n - 10$ (Theorem 1)	$M \leq 6.5n - 13^*$ (Theorem 11)	$M \leq 74.2n^*$ (Theorem 12)	$M = \frac{n(n-1)}{2}$ (Theorem 10)
ACEα drawing	$M \leq 3(3n-6)^*$ (Theorem 17)	$M \leq 27n^*$ (Theorem 18)	$M \leq 477n^*$ (Theorem 19)	$M = \frac{n(n-1)}{2}$ (Theorem 10)
ACLα drawing	$M \leq \frac{\pi}{\alpha}(3n-6)^*$ (Theorem 16)	$M = \frac{n(n-1)}{2}$ (Theorem 20)	$M = \frac{n(n-1)}{2}$ (Theorem 20)	$M = \frac{n(n-1)}{2}$ (Theorem 20)

The symbol "*" indicates that it is not known whether the bound is tight.

geometric graph theory and graph drawing. We conclude this survey by listing some research directions that, in our opinion, are among the most promising. We shall distinguish among Turán-type problems, recognition and characterization problems, optimization problems, and algorithm engineering problems.

5.1 Turán-Type Problems

Table 1 summarizes the Turán-type results described in this survey chapter. In the table, M denotes the maximum number of edges in a drawing of a specific type. Every entry of the table contains an upper bound to the value M for a different type of drawing (row), with a specific curve complexity (column). Not all bounds in Table 1 are known to be tight. A natural question is therefore the following.

Open Problem 1. *Improve the upper bounds for those entries marked with the symbol "*" in Table 1 or show that these bounds are tight.*

5.2 Recognition and Characterization Problems

As stated in Theorem 3, recognizing those graphs that can be realized as geometric RAC graphs is $\mathcal{NP}$-hard.

Open Problem 2. *What is the complexity of recognizing graphs that admit a geometric ACEα drawing or a geometric ACLα drawing, for a given angle $\alpha \in (0, \frac{\pi}{2})$?*

By Theorem 10, every graph admits an RAC drawing with at most three bends per edge. Also, by Theorem 20, every graph admits an ACLα drawing with at most one bend per edge, for a given $\alpha \in (0, \frac{\pi}{2})$.

Open Problem 3. *What is the complexity of recognizing graphs that admit an RAC drawing with at most one bend per edge? And for those having an RAC drawing with at most two bends per edge?*

Open Problem 4. *What is the complexity of recognizing graphs that admit an ACEα drawing with at most one or at most two bends per edge, for a given $\alpha \in (0, \frac{\pi}{2})$?*

Strictly related problems consider the characterization of families of graphs that admit drawings with good crossing-angle resolution. For example, one can study the following extremal problem (Theorem 2 may be a starting point for studying this problem).

Open Problem 5. *What graphs having exactly $4n - 10$ edges (i.e., of maximum edge density) can be realized as geometric RAC graphs?*

Another characterization problem considers low-degree graphs. Duncan et al. [26] prove that graphs with maximum vertex degree 4 have geometric thickness 2. As explained in Sect. 2, straight-line RAC drawings also have geometric thickness 2. However, an immediate consequence of Theorem 7 is that not all graphs with maximum vertex degree 4 can be realized as geometric RAC graphs (for example, $K_{4,4}$ cannot be a geometric RAC graph). Note that, by Theorem 14, any graph with maximum vertex degree 3 admits an RAC drawing with curve complexity 1.

Open Problem 6. *Can all graphs with maximum vertex degree 3 be realized as geometric RAC graphs?*

5.3 Optimization Problems

Many of the typical optimization questions that have been asked in the literature for planar drawings can also be asked for drawings with good crossing-angle resolution. For example, it is well known that a quadratic size grid is always sufficient and sometimes necessary for straight-line planar drawings [12]. On the other hand, if edge crossings are allowed, then every planar graph admits a drawing in linear area [51]. Theorem 9 proves that $\Omega(n^2)$ area may also be required by geometric RAC graphs.

Open Problem 7. *What is the area requirement of polyline RAC drawings of planar graphs? What about the area requirement of polyline ACEα or ACLα drawings of planar graphs?*

The above open problem can be made more specific by looking at subfamilies of planar graphs. For example, Frati [30] shows an $\Omega(n2^{\sqrt{\log n}})$ lower bound to the area requirement of polyline drawings of series-parallel graphs.

Open Problem 8. *Can one compute polyline RAC drawings (or ACEα or ACLα drawings) of series-parallel graphs with low curve complexity and area $o(n2^{\sqrt{\log n}})$?*

Finally, Theorem 15 establishes a tradeoff between the area and curve complexity of RAC drawings of nonplanar graphs.

Open Problem 9. *Is it possible to compute RAC drawings of graphs in $o(n^3)$ area and curve complexity at most 4?*

Another natural research direction is to investigate the problem of computing drawings such that the crossing-angle resolution is maximized. More precisely, we ask the following.

Open Problem 10. *Compute straight-line drawings that optimize one of the following functions:*

- *The crossing-angle resolution.*
- *The sum of all crossing angles.*
- *The average value over all crossing angles.*
- *The total angle resolution, the minimum between crossing- and vertex-angle resolution.*

Results related to Open Problem 10 can be found in [6, 14, 15]. In [14, 15], upper and lower bounds on the crossing-angle resolution of straight-line drawings of complete graphs are presented; in [6], asymptotically optimal bounds on the total angle resolution of circular drawings of complete graphs and of two-layer drawings of complete bipartite graphs are presented.

5.4 Algorithm Engineering Problems

The need to visually analyze large relational datasets justifies a significant effort into the technology transfer of algorithmic solutions for drawing nonplanar graphs. In this context, the development of libraries and experimental platforms devoted to algorithms that compute drawings with good crossing angle and total angle resolution is an interesting research direction. Different heuristics have been designed and experimentally tested that perform well in terms of crossing-angle or total angle resolution. Extensions of classical force-directed algorithms that take into account the optimization of the crossing-angle resolution are described in [22, 34]. The algorithm in [22] allows bent edges, and it is used to compute simultaneous drawings of graphs of concepts in a system for the visual analysis of web-traffic data. A force-directed algorithm that attempts to optimize the total angle resolution is described in [6]. Experimental evaluations of classical and enhanced force-directed algorithms that take into account different aesthetic criteria, including crossing- and vertex-angle resolutions, are presented in [34]. A graph drawing framework combining the topology-shape-metrics and the force-directed approaches with the goal of computing drawings with good tradeoffs among the number of crossings, vertex-angle resolution, crossing-angle resolution, and geodesic edge tendency is described and experimentally evaluated in [23]. A postprocessing algorithm, based on quadratic programming, that improves the crossing-angle resolution of a circular layout is described in [39].

Besides force-directed techniques, it would be interesting to design and experimentally evaluate other algorithmic approaches. For example, Theorem 6 motivates the following open problem, which can be used as a building block to implement a variant of the well-known Sugiyama algorithmic framework for layered graph drawing (see, e.g., [46]).

Open Problem 11. *Design and experimentally evaluate heuristics and/or approximation algorithms for extracting the maximum two-layer RAC drawable subgraph of a given bipartite graph.*

Acknowledgements Our work is supported in part by MIUR of Italy under project AlgoDEEP prot. 2008TFBWL4.

References

1. E. Ackerman, On the maximum number of edges in topological graphs with no four pairwise crossing edges. Discr. Comput. Geom. **41**(3), 365–375 (2009)
2. E. Ackerman, R. Fulek, C.D. Tóth, On the size of graphs that admit polyline drawings with few bends and crossing angles, in *Proceedings of GD 2010*. LNCS, vol. 6502 (Springer, Berlin, 2010), pp. 1–12
3. E. Ackerman, G. Tardos, On the maximum number of edges in quasi-planar graphs. J. Comb. Theor. Ser. A **114**(3), 563–571 (2007)
4. P. Angelini, L. Cittadini, G. Di Battista, W. Didimo, F. Frati, M. Kaufmann, A. Symvonis, On the perspectives opened by right angle crossing drawings, in *Proceedings of GD 2009*. LNCS, vol. 5849 (Springer, Berlin, 2010), pp. 21–32
5. P. Angelini, L. Cittadini, G. Di Battista, W. Didimo, F. Frati, M. Kaufmann, A. Symvonis, On the perspectives opened by right angle crossing drawings. J. Graph Algorithm. Appl. **15**(1), 53–78 (2011) (Special issue on GD 2009)
6. E. Argyriou, M. Bekos, A. Symvonis, Maximizing the total resolution of graphs, in *Proceedings of GD 2010*. LNCS, vol. 6502 (Springer, Berlin, 2010), pp. 62–67
7. E.N. Argyriou, M.A. Bekos, A. Symvonis, The straight-line RAC drawing problem is NP-hard, in *SOFSEM*. Lecture Notes in Computer Science, vol. 6543 (Springer, Berlin, 2011), pp. 74–85
8. K. Arikushi, R. Fulek, B. Keszegh, F. Moric, Csaba D. Tóth, Graphs that admit right angle crossing drawings. Comput. Geom. Springer, Berlin, **45**(4), 169–177 (2012)
9. C. Berge, *Graphs* (North Holland, Amsterdam, 1885)
10. O.V. Borodin, A.V. Kostochka, A. Raspaud, E. Sopena, Acyclic colouring of 1-planar graphs. Discr. Appl. Math. **114**(1–3), 29–41 (2001)
11. G. Brightwell, E.R. Scheinerman, Representations of planar graphs. SIAM J. Discr. Math. **6**(2), 214–229 (1993)
12. H. de Fraysseix, J. Pach, R. Pollack, How to draw a planar graph on a grid. Combinatorica **10**(1), 41–51 (1990)
13. G. Di Battista, P. Eades, R. Tamassia, I.G. Tollis, *Graph Drawing* (Prentice Hall, Upper Saddle River, NJ, 1999)
14. E. Di Giacomo, W. Didimo, P. Eades, S.-H. Hong, G. Liotta, On the crossing resolution of complete geometric graphs. Technical report, DIEI RT-002-11, University of Perugia, Italy (2011). http://www.diei.unipg.it/rt/rt_frames.htm
15. E. Di Giacomo, W. Didimo, P. Eades, S.-H. Hong, G. Liotta, Bounds on the crossing resolution of complete geometric graphs. Discr. Appl. Math., **160**(1–2), 132–139 (2012). Available online, doi:10.1016/j.dam.2011.09.016

16. E. Di Giacomo, W. Didimo, P. Eades, G. Liotta, 2-layer right angle crossing drawings, in *IWOCA 2011*, Springer, Berlin, LNCS, **7056**, 156–169 (2011)
17. E. Di Giacomo, W. Didimo, G. Liotta, H. Meijer, Area, curve complexity, and crossing resolution of non-planar graph drawings, in *Proceedings of GD 2009*. LNCS, vol. 5849 (Springer, Berlin, 2010), pp. 15–20
18. E. Di Giacomo, W. Didimo, G. Liotta, H. Meijer, Area, curve complexity, and crossing resolution of non-planar graph drawings. Theor. Comput. Syst. **49**(3), 565–575 (2011)
19. W. Didimo, P. Eades, G. Liotta, Drawing graphs with right angle crossings, in *Proceedings of WADS 2009*. LNCS, vol. 5664 (Springer, Berlin, 2009), pp. 206–217
20. W. Didimo, P. Eades, G. Liotta, A characterization of complete bipartite rac graphs. Inform. Process. Lett. **110**(16), 687–691 (2010)
21. W. Didimo, P. Eades, G. Liotta, Drawing graphs with right angle crossings. Theor. Comput. Sci. **412**(39), 5156–5166 (2011)
22. W. Didimo, G. Liotta, S.A. Romeo, Graph visualization techniques for conceptual web site traffic analysis, in *PacificVis*, IEEE, 2010, pp. 193–200
23. W. Didimo, G. Liotta, S.A. Romeo, Topology-driven force-directed algorithms, in *Proceedings of GD 2010*. LNCS, vol. 6502 (Springer, Berlin, 2010), pp. 165–176
24. V. Dujmovic, J. Gudmundsson, P. Morin, T. Wolle, Notes on large angle crossing graphs. CoRR, abs/0908.3545 (2009). http://arxiv.org/abs/0908.3545
25. V. Dujmovic, J. Gudmundsson, P. Morin, T. Wolle, Notes on large angle crossing graphs. Chicago J. Theor. Comput. Sci., Brisbane, Australia, January 18–21 (2011)
26. C.A. Duncan, D. Eppstein, S.G. Kobourov, The geometric thickness of low degree graphs, in *Symposium on Computational Geometry*, ACM, 2004, pp. 340–346
27. P. Eades, G. Liotta, Right angle crossing graphs and 1-planarity, in *proceedings of the 27th European Workshop on Computational Geometry (EuroCG)*, March 28–30, Morschach, Switzerland, pp. 155–158 (2011)
28. P. Eades, A. Symvonis, S. Whitesides, Three-dimensional orthogonal graph drawing algorithms. Discr. Appl. Math. **103**(1–3), 55–87 (2000)
29. I. Fabrici, T. Madaras, The structure of 1-planar graphs. Discr. Math. **307**(7–8), 854–865 (2007)
30. F. Frati, Improved lower bounds on the area requirements of series-parallel graphs, in *Proceedings of GD 2010*. LNCS, vol. 6502 (Springer, Berlin, 2010), pp. 220–225
31. A. Garg, R. Tamassia, On the computational complexity of upward and rectilinear planarity testing. SIAM J. Comput. **31**(2), 601–625 (2001)
32. W. Huang, Using eye tracking to investigate graph layout effects, in *APVIS 2007*, 6th International Asia-Pacific Symposium on Visualization, Sydney, Australia, 5–7 February 2007, IEEE, pp. 97–100 (2007)
33. W. Huang, An eye tracking study into the effects of graph layout. CoRR, abs/0810.4431 (2008). http://arxiv.org/abs/0810.4431
34. W. Huang, P. Eades, S.-H. Hong, C.-C. Lin, Improving force-directed graph drawings by making compromises between aesthetics, in *VL/HCC*, IEEE, 2010, pp. 176–183
35. W. Huang, S.-H. Hong, P. Eades, Effects of crossing angles, in IEEE VGTC Pacific Visualization Symposium 2008, PacificVis 2008, Kyoto, Japan, March 5–7, pp. 41–46 (2008)
36. M. Jünger, P. Mutzel (eds.), *Graph Drawing Software (Mathematics and Visualization)* (Springer-Verlag, Berlin, 2003)
37. M. Kaufmann, D. Wagner (eds.), *Drawing Graphs, Methods and Models (the Book Grow Out of a Dagstuhl Seminar, April 1999)*. Lecture Notes in Computer Science, vol. 2025 (Springer-Verlag, Berlin, 2001)
38. V.P. Korzhik, B. Mohar, Minimal obstructions, Springer, Berlin, Lecture Notes in Computer Science, **5417**, 302–312 (2008)
39. Q. Nguyen, P. Eades, S.-H. Hong, W. Huang, Large crossing angles in circular layouts, in *Proceedings of GD 2010*. LNCS, vol. 6502 (Springer, Berlin, 2010), pp. 397–399
40. J.E. Goodman, J. O'Rourke (eds.), Geometric graph theory, in *Handbook of Discrete and Computational Geometry* (CRC, 2004), Chapman and Hall/CRC, New York, United States of America, pp. 219–238

41. J. Pach, R. Radoičić, G. Tardos, G. Tóth, Improving the crossing lemma by finding more crossings in sparse graphs. Discr. Comput. Geom. **36**(4), 527–552 (2006)
42. J. Pach, G. Tóth, Graphs drawn with few crossings per edge. Combinatorica **17**(3), 427–439 (1997)
43. J. Petersen, Die theorie der regulären graphen. Acta Math. **15**, 193–220 (1891)
44. H.C. Purchase, Effective information visualisation: a study of graph drawing aesthetics and algorithms. Interact. Comput. **13**(2), 147–162 (2000)
45. H.C. Purchase, D.A. Carrington, J.-A. Allder, Empirical evaluation of aesthetics-based graph layout. Empir. Softw. Eng. **7**(3), 233–255 (2002)
46. K. Sugiyama, *Graph Drawing and Applications for Software and Knowledge Engineers* (World Scientific, Singapore, 2002)
47. Y. Suzuki, Optimal 1-planar graphs which triangulate other surfaces. Discr. Math. **310**(1), 6–11 (2010)
48. J.J. Thomas, K.A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics* (National Visualization and Analytics Ctr, 2005), Pacific Northwest National Laboratory, United States of America
49. M. van Kreveld, The quality ratio of RAC drawings and planar drawings of planar graphs, in *Proceedings of GD 2010*. LNCS, vol. 6502 (Springer, Berlin, 2010), pp. 371–376
50. C. Ware, H.C. Purchase, L. Colpoys, M. McGill, Cognitive measurements of graph aesthetics. Inform. Vis. **1**(2), 103–110 (2002)
51. D.R. Wood, Grid drawings of *k*-colourable graphs. Comput. Geom. **30**(1), 25–28 (2005)

Mover Problems

Adrian Dumitrescu

Abstract Leo Moser asked what is the region of largest area that can be moved around a right-angled corner in a corridor of unit width? Similarly, what is the region of largest area that can be reversed in a T junction made up of roads of unit width? Can a specific 3-dimensional object pass through a given door? Our survey aims at showing that mover problems are no less challenging even if there are no obstacles other than the objects themselves, but there are many objects to move. We survey some recent results on motion planning and reconfiguration for systems of multiple objects and for modular systems with applications in robotics, and collect some open problems coming out of this line of research.

1 Introduction

The motion planning problem in its simplest form is that of finding a collision-free movement for a single object from a given start position to another specified target position in the presence of obstacles; see, e.g., [34]. Besides the pure algorithmic aspect, a natural question is how large such an object can be moved through a given corridor formed by the obstacles. For instance, Leo Moser asked for the region of largest area that can be moved around a right-angled corner in a corridor of unit width [36]; see also [19, G5]. The problem became known as the *piano mover's problem*, or more accurately (since any single mover would be overwhelmed by the task), the *piano movers' problem*.

The motion planning problem can also be formulated for systems of multiple independent objects. Schwartz and Sharir present an algorithm that solves the following motion planning problem that arises in robotics: Given a start and a

Supported in part by NSF CAREER grant CCF-0444188 and NSF grant DMS-1001667.

A. Dumitrescu (✉)
Department of Computer Science, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
e-mail: dumitres@uwm.edu.

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_11,

target configuration of n disks, and a region bounded by a collection of walls, find a continuous motion connecting the two configurations of these bodies during which they avoid collision with the walls and with each other, or establish that no such motion exists [46]. Motion planning for multiple objects has been also addressed in [28]. Again, besides the pure algorithmic aspect, a natural question is to estimate the number of moves (of a certain type) such a reconfiguration requires in terms of n, the number of objects. In this survey we are primarily concerned with this combinatorial aspect of moving multiple objects. We discuss three related reconfiguration scenarios for systems of multiple objects. The reader is also referred to the two surveys on related topics by Pach and Sharir [39, Chap. 9] and by Demaine and Hearn [20].

1. A *body* (or *object*) in the plane is a compact connected set in $\mathbb{R}^2$ with a nonempty interior. Two initially disjoint bodies *collide* if they share an interior point at some time during their motion. Consider a set of n pairwise interior-disjoint objects in the plane that need to be brought from a given start (initial) configuration S into a desired target (goal) configuration T, without causing collisions. The *reconfiguration* problem for such a system is that of computing a sequence of object motions (a schedule, or motion plan) that achieves this task. Depending on the existence of such a sequence of motions, we call that instance of the problem *feasible* and, respectively, *infeasible*.

   Our reconfiguration problem is a simplified version of the multirobot motion planning problem [32], in which a system of robots are operating together in a shared workplace and from time to time need to move from their initial positions to a set of target positions. The workspace is often assumed to extend throughout the entire plane, with no obstacles other than the robots themselves. In many applications, the robots are indistinguishable so any of them can occupy any of the specified target positions. Another application that permits the same abstraction is moving around large sets of heavy objects in a warehouse. Typically, one is interested in minimizing the number of moves and in designing efficient algorithms for carrying out the motion plan. There are several types of moves, such as sliding, translation, or lifting, which lead to three different models that will be discussed in Sect. 2.
2. A different kind of reconfiguration problem appears in connection to *metamorphic* or *self-reconfigurable* modular systems. A *modular* robotic system consists of a number of identical robotic modules that can connect to, disconnect from, and relocate relative to adjacent modules; see examples in [14, 18, 37, 38, 40, 45, 48–50, 53]. Typically, the modules have a regular symmetry so that they can be packed densely, with small gaps between them. Such a system can be viewed as a large swarm of physically connected robotic modules that behave collectively as a single entity.

   The system changes its overall shape and functionality by reconfiguring into different formations. In most cases individual modules are not capable of moving by themselves; however, the entire system may be able to move to a new position when its members repeatedly change their positions relative to their neighbors, by rotating or sliding around other modules [13, 37, 52], or by expansion and contraction [45]. In this way the entire system, by changing its aggregate geometric structure, may acquire new functionalities to accomplish a given task or to interact with the environment.

Shape changing in these composite systems is envisioned as a means to accomplish various tasks, such as reconnaissance, exploration, satellite recovery, or operation in constrained environments inaccessible to humans (e.g., nuclear reactors, space, or deep water). For another example, a self-reconfigurable robot can aggregate as a snake to traverse a tunnel and then reconfigure as a six-legged spider to move over uneven terrain. A novel useful application is to realize self-repair: A self-reconfigurable robot carrying some additional modules may abandon the failed modules and replace them with spare units [45]. It is usually assumed that the modules must remain connected all (or most) of the time during reconfiguration.

The motion planning problem for such a system is that of computing a sequence of module motions that brings the system in a given initial configuration I into a desired goal configuration G. Reconfiguration for modular systems acting in a grid-like environment, and where moves must maintain connectivity of the whole system has been studied in [25–27], focusing on two basic capabilities of such systems: reconfiguration and locomotion. We present details in Sect. 3.

3. In some cases the problem of bringing a set of pairwise disjoint objects (in the plane or in the space) to a desired goal, configuration, admits the following abstraction: We have an underlying finite or infinite connected graph; the start configuration is represented by a set of n chips at n start vertices and the target configuration by another set of n target vertices. A vertex can be both a start and a target position. The case when the chips are labeled or unlabeled gives two different variants of the problem. In one move a chip can follow an arbitrary path in the graph and end up at another vertex, provided the path (including the end vertex) is free of other chips [16].

The motion planning problem for such a system is that computing a sequence of chip motions that brings the chips from their initial positions to their target positions. Again, the problem may be feasible or infeasible. We address multiple aspects of this variant in Sect. 4. We note that the three models mentioned earlier in (1) do not fall in the above graph reconfiguration framework, because an object may partially overlap several target positions.

2 Models of Reconfiguration for Systems of Objects in the Plane

We formulate these models for systems of disks, since they are simpler and most of our results are for this class of convex bodies. These rules can be extended (not necessarily uniquely) for arbitrary convex bodies in the plane. The decision problems we refer to below, pertaining to various reconfiguration problems we discuss here, are in standard form and concern systems of (arbitrary or congruent) disks. For instance, the *reconfiguration problem* U-SLIDE-RP for congruent disks is as follows: Given a start configuration and a target configuration, each with n unlabeled congruent disks in the plane, and a positive integer k, is there a reconfiguration motion plan with at most k sliding moves? It is worth clarifying that for the unlabeled variant, if the start and target configuration contain subsets of congruent disks, there is freedom is choosing which disks will occupy target positions. However, in the labeled variant, this assignment is uniquely determined by the labeling; of course, a valid labeling must respect the size of the disks.

1. *Sliding model*: One move is sliding a disk to another location in the plane without colliding with any other disk, where the disk center moves along an arbitrary continuous curve. This model was introduced in [8]. The labeled and unlabeled variants are L-SLIDE-RP and U-SLIDE-RP, respectively.
2. *Translation model*: One move is translating a disk to another location in the plane along a fixed direction without colliding with any other disk. This is a restriction imposed to the sliding model above for making each move as simple as possible. This model was introduced in [2]. The labeled and unlabeled variants are L-TRANS-RP and U-TRANS-RP, respectively.
3. *Lifting model*: One move is lifting a disk and placing it back in the plane anywhere in the free space, that is, at a location where it does not intersect (the interior of) any other disk. This model was introduced in [7]. The labeled and unlabeled variants are L-LIFT-RP and U-LIFT-RP, respectively.

It turns out that moving a set of objects from one place to another is related to certain *separability* problems [12, 17, 29, 31]; see also [44]. For instance, given a set of disjoint polygons in the plane, may each be moved "to infinity" in a continuous motion in the plane without colliding with the others? Often constraints are imposed on the types of motions allowed, e.g., only translations, or only translations in a fixed set of directions. Usually only one object is permitted to move at a time. Without the convexity assumption on the objects, it is easy to show examples when the objects are interlocked and could only be moved "together" in the plane; however, they could be easily separated using the third dimension, i.e., in the lifting model.

It can be shown that for the class of disks, the reconfiguration problem in each of these models is always feasible [2, 7, 8, 12, 29, 31]. This follows essentially from the feasibility in the sliding model and the translation model; see Sect. 2.1. For the more general class of convex objects, one needs to allow rotations. For simplicity, we restrict ourselves mostly to the case of disks. We are thus led to the following generic question: Given a pair of start and target configurations, each consisting of n pairwise disjoint disks in the plane, what is the minimum number of moves that suffice for transforming the start configuration into the target configuration for each of these models?

If no target disk coincides with a start disk, so each disk must move at least once, obviously at least n moves are required. In general, one can use (a variant of) the following simple *universal* algorithm for the reconfiguration of n objects

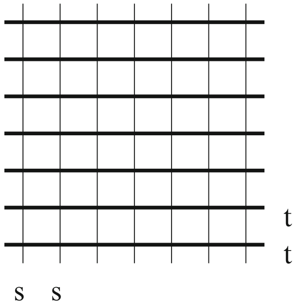


Fig. 1 $2n-1$ moves are necessary (in the sliding or the lifting model) to bring the n segments from *vertical position* to *horizontal position*

Table 1 Comparison summary: number of moves for disks in the plane/chips in the grid

Model	Type	Lower bound	Upper bound
Translating disks	Congruent (unlabeled)	$\lfloor 5n/3 \rfloor - 1$	$2n-1$
	Arbitrary (unlabeled)	$2n$	$2n$
Sliding disks	Congruent (unlabeled)	$16n/15 - o(n)$	$3n/2 + o(n)$
	Arbitrary (unlabeled)	$2n - o(n)$	$2n-1$
Lifting disks	Congruent (unlabeled)	$n + \Omega(n^{1/2})$	$n + O(n^{2/3})$
	Arbitrary (unlabeled)	$\lfloor 5n/3 \rfloor$	$9n/5$
Sliding chips (grid)	Unlabeled	n	n
	Labeled	$3n/2$	$7n/4$
Lifting chips (grid)	Unlabeled	n	n
	Labeled	$3n/2$	$3n/2$

using $2n$ moves. To be specific, consider the lifting model. In the first step (n moves), move all the objects away anywhere in the free space. In the second step (n moves), bring the objects "back" to target positions. For the class of segments (or rectangles) as objects, it is easy to construct examples that require $2n-1$ moves for reconfiguration, in any of the three models, even for congruent objects; see Fig. 1. A first goal is to estimate more precisely where in the interval $[n, 2n]$ the answer lies for each of these models. The best current lower and upper bounds on the number of moves necessary in the three models described earlier are listed in Table 1. It is quite interesting to compare the bounds on the number of moves for the three models, translation, sliding, and lifting, with those for the graph variants discussed in Sect. 4. Table 1, which is constructed on the basis of the results in [2, 7, 8, 16], facilitates this comparison.

Some remarks are in order. Clearly, any lower bound (on the number of moves) for lifting is also valid for sliding, and any upper bound (on the number of moves) for sliding is also valid for lifting. Another observation is that for lifting, those objects whose target position coincides with their start position can be safely ignored, while for sliding this is not true. A simple example is illustrated in Fig. 2: Assume that we have a large disk surrounded by $n-1$ smaller ones. The large disk has to be moved to another location, while the $n-1$ smaller disks have to stay where they are.

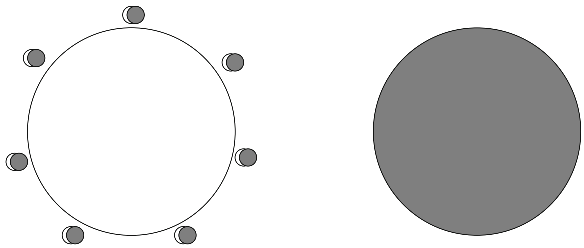


Fig. 2 One move is enough in the lifting model, while $n-1$ are needed in the sliding model for this pair of start and target configurations with n disks each (here $n = 8$). Start disks are *white* and target disks are *shaded*

One move is enough in the lifting model, while $n-1$ are needed in the sliding model: One needs to make space for the large disk to move out by moving out about half of the small disks and then moving them back in to the same positions.

A move is a *target move* if it moves a disk to a final target position. Otherwise, it is a *nontarget* move. Our lower bounds use the following argument: If no target disk coincides with a start disk (so each disk must move), a schedule that requires x nontarget moves must consist of at least $n+x$ moves.

2.1 The Sliding Model

It is not difficult to show that, for the class of disks, the reconfiguration problem in the sliding model is always feasible. More generally, the problem remains feasible for the class of all convex objects using sliding moves; this follows from Theorem 1. This old result appears in the work of Fejes Tóth and Heppes [29], but it can be traced back to de Bruijn [12]; some algorithmic aspects of the problem have been addressed subsequently by Guibas and Yao [31].

Theorem 1 ([12,29,31]). *Any set of n convex objects in the plane can be separated via translations all parallel to any given fixed direction, with each object moving once only. If the topmost and bottommost points of each object are given (or can be computed in $O(n\log n)$ time), an ordering of the moves can be computed in $O(n\log n)$ time.*

The universal algorithm mentioned earlier can be adapted to perform th reconfiguration of any set of n convex objects. It performs $2n$ moves for the reconfiguration of n disks. In the first step (n moves), in decreasing order of the x-coordinates of their centers, slide the disks initially along a horizontal direction, one by one to the far right. Observe that no collisions can occur. In the second step (n moves), bring the disks "back" to target positions in increasing order of the x-coordinates of their centers. (General convex objects may need rotations and translations in the second step.) Already for the class of disks, Theorem 3 shows that one cannot do much better in terms of the number of moves. The following bounds on the number of moves for translating disks are due to Bereg et al. [8] (Theorems 2 and 3).

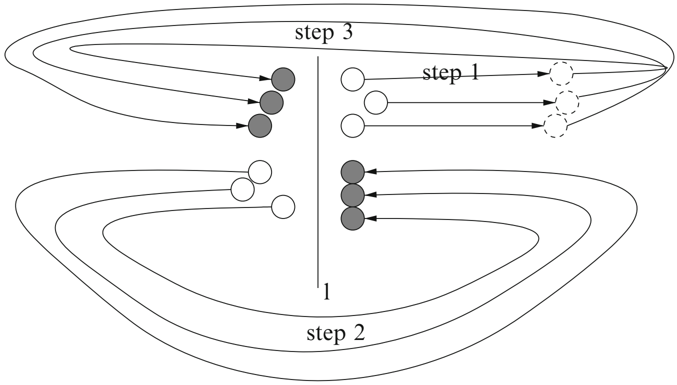


Fig. 3 Algorithm with three steps for sliding congruent disks. The start disks are *white* and the target disks are *shaded*

Theorem 2 ([8]). *Given a pair of start and target configurations S and T, each consisting of n congruent disks,* $\frac{3n}{2} + O(\sqrt{n \log n})$ *sliding moves always suffice for transforming the start configuration into the target configuration. The entire motion can be computed in* $O(n^{3/2}(\log n)^{-1/2})$ *time. On the other hand, there exist pairs of configurations that require* $\left(1 + \frac{1}{15}\right) n - O(\sqrt{n})$ *moves for this task.*

We now briefly sketch the upper bound proof and the corresponding algorithm in [8] for congruent disks. First, one shows the existence of a line bisecting the set of centers of the start disks such that the strip of width 6 around this line contains a small number of disks. More precisely, the following holds:

Lemma 2.1 ([8]). *Let S be a set of n pairwise disjoint unit (radius) disks in the plane. Then there exists a line* ℓ *that bisects the centers of the disks such that the parallel strip of width* 6 *around* ℓ *(that is,* ℓ *runs in the middle of this strip) contains entirely at most* $O(\sqrt{n \log n})$ *disks.*

Let S' and T' be the centers of the start disks and target disks, respectively, and let ℓ be the line guaranteed by Lemma 2.1. We can assume that ℓ is vertical. Denote by $s_1 = \lfloor n/2 \rfloor$ and $s_2 = \lceil n/2 \rceil$ the number of centers of start disks to the left and right of ℓ (centers on ℓ can be assigned to the left or right). Let $m = O(\sqrt{n \log n})$ be the number of start disks contained entirely in the vertical strip of width 6 around ℓ. Denote by t_1 and t_2 the number of centers of target disks to the left and right of ℓ, respectively. By symmetry, we can assume that $t_1 \le n/2 \le t_2$.

Let R be a region containing all start and target disks, e.g., the smallest axis-aligned rectangle that contains all disks. The algorithm has three steps. All moves in the region R are taken along horizontal lines, i.e., perpendicularly to the line ℓ. The reconfiguration procedure is schematically shown in Fig. 3. This illustration ignores the disks/targets in the center parallel strip.

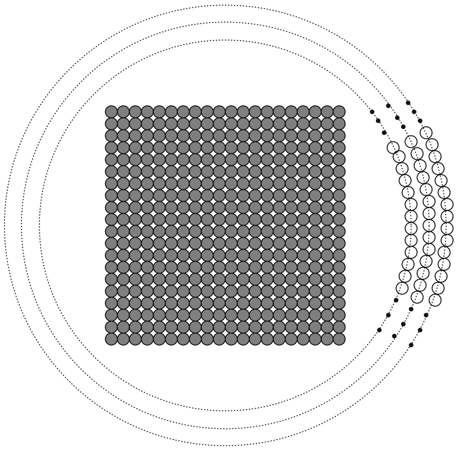


Fig. 4 A lower bound of $n+\Omega(\sqrt{n})$ moves: $\Theta(\sqrt{n})$ rings, each with $\Theta(\sqrt{n})$ start disk positions. Targets are densely packed in a square formation enclosed by the rings

Step 1 Slide to the far right all start disks whose centers are to the right of ℓ and the (other) start disks in the strip, one by one, in decreasing order of their x-coordinates (with ties broken arbitrarily). At this point all $t_2 \geq n/2$ target disk positions whose centers lie right of ℓ are free.

Step 2 In increasing order of their x-coordinates, fill free target positions whose centers are right of ℓ using all the $s_1' \leq s_1 \leq n/2$ remaining disks whose centers are to the left of ℓ. These disks are taken in increasing order of their x-coordinates: Each disk translates first to the left, then to the right on a wide arc, and to the left again in the end. Note that $s_1' \leq n/2 \leq t_2$. Now all the target positions whose centers lie left of ℓ are free.

Step 3 Move to place the far away disks: First continue to fill target positions whose centers are to the right of ℓ, in increasing order of their x-coordinates. When done, fill target positions whose centers are left of ℓ, in decreasing order of their x-coordinates. Note that at this point all target positions whose centers lie left of ℓ are free.

The only nontarget moves are those done in step 1 and their number is $n/2 + O(\sqrt{n \log n})$, so the total number of moves is $3n/2 + O(\sqrt{n \log n})$.

A first idea in constructing a lower bound is as follows: The target configuration consists of a set of n densely packed unit (radius) disks contained, for example, in a square of side length $\approx 2\sqrt{n}$. The disks in the start configuration enclose the target positions using concentric rings, that is, $\Theta(\sqrt{n})$ rings, each with $\Theta(\sqrt{n})$ start disk positions, as shown in Fig. 4. Now observe that for each ring, the first move that involves a disk in that ring must be a nontarget move. The number of rings is $\Theta(\sqrt{n})$, from which a lower bound of $n + \Omega(\sqrt{n})$ follows.

This basic idea of a cage-like construction can be further refined by redesigning the cage [8]. The new design is more complicated and uses "rigidity" considerations, which go back to stable disk packings of density 0 due to Böröczky [9]. A packing $\mathcal{C}$ of unit (radius) disks in the plane is said to be *stable* if each disk is kept fixed

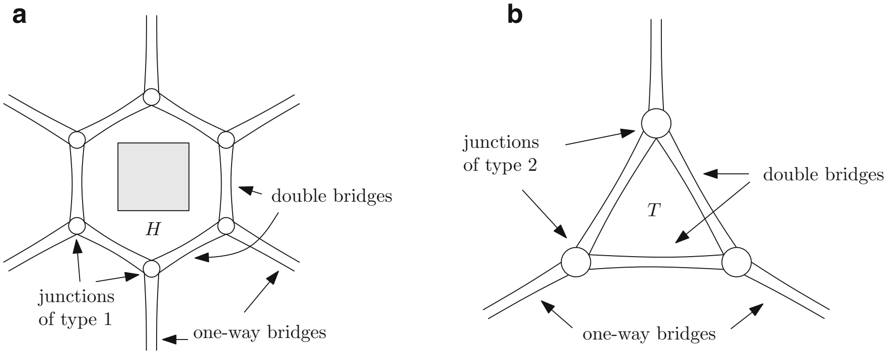


Fig. 5 Two start configurations based on hexagonal and triangular cage-like constructions. Targets are densely packed in a square formation enclosed by the cage, as shown in (**a**); those in (**b**) are labeled "T"

by its neighbors [11]. More precisely, $\mathcal{C}$ is stable if none of its elements can be translated by any small distance in any direction without colliding with the others. It is easy to see that any stable system of (unit) disks in the plane must have infinitely many elements. Böröczky [9] showed that somewhat surprisingly, there exist stable systems of unit disks with an arbitrarily small density (i.e., the total area of the disks inside any sufficiently large axis-aligned square is arbitrarily small). These systems can be adapted for the purpose of constructing a lower bound in the sliding model for congruent disks. The details are quite technical, and we only sketch here the new cage-like constructions depicted in Fig. 5.

Let's refer to the disks in the start (respectively target) configuration as white (respectively, black) disks. Now fix a large n, and take n white disks. Use $O(\sqrt{n})$ of them to build three junctions connected by three "double-bridges" to enclose a triangular region that can accommodate n tightly packed nonoverlapping black disks, as shown in Fig. 5b. Divide the remaining white disks into three roughly equal groups, each of size $\frac{n}{3} - O(\sqrt{n})$, and rearrange each group to form the initial section of a "one-way bridge" attached to the unused sides of the junctions. Each of these bridges consists of five rows of disks of "length" roughly $\frac{n}{15}$, where the length of a bridge is the number of disks along its side. The design of both the junctions and the bridges prohibits any target move before one moves out a sequence of about $\frac{1}{5} \cdot \frac{n}{3} = \frac{n}{15}$ white adjacent disks starting at the far end of one of the one-way bridges. The reason is that with the exception of the at most $3 \times 4 = 12$ white disks at the far ends of the truncated one-way bridges, every white disk is fixed by its neighbors. The total number of necessary moves is at least $\left(1 + \frac{1}{15}\right) n - O(\sqrt{n})$ for this triangular ring construction, and at least $\left(1 + \frac{1}{30}\right) n - O(\sqrt{n})$ for the hexagonal ring construction. Observe that the triangular cage yields a better bound.

For disks of arbitrary radii, the following result is obtained by the same authors [8]:

Theorem 3 ([8]). *Given a pair of start and target configurations, each consisting of n disks of arbitrary radii,* 2*n sliding moves always suffice for transforming the*

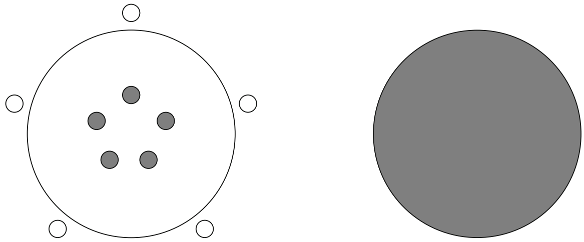


Fig. 6 A simple configuration that requires about $3n/2$ moves (basic step for the recursive construction)

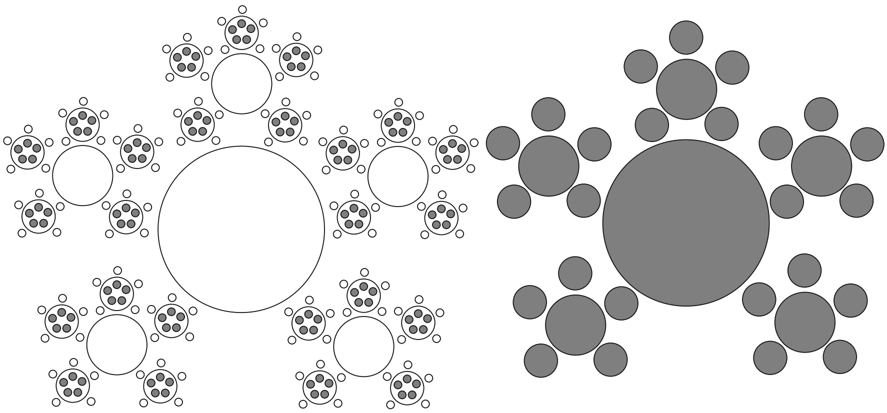


Fig. 7 Recursive lower-bound construction for sliding disks of arbitrary radii: $m = 2$ and $k = 3$

start configuration into the target configuration. The entire motion can be computed in $O(n \log n)$ time. On the other hand, there exist pairs of configurations that require $2n - o(n)$ moves for this task.

The upper bound follows from the universal reconfiguration algorithm described earlier. The lower bound is a recursive construction shown in Fig. 7. It is obtained by iterating recursively the basic construction in Fig. 6, which requires about $3n/2$ moves: Note that the target positions of the $n-1$ small disks lie inside the start position of the large disk. This means that no small disk can reach its target before the large disk moves away, that is, before roughly half of the $n-1$ small disks move away. So about $n/2$ nontarget moves are necessary; thus, about $3n/2$ moves in total are necessary.

In the recursive construction, the small disks around a large one are replaced by the "same" construction scaled (see Fig. 7). All disks have distinct radii, so it may be convenient to think of them as being labeled. There are one large disk and $2m+1$ groups of smaller disks around it close to the vertices of a regular $(2m+1)$-gon ($m \geq 1$). The small disks on the last level or recursion have targets inside the big ones they surround (the other disks have targets somewhere else). Let $m \geq 1$ be fixed. If there are k levels in the recursion, then instead of about $n/2$ nontarget moves

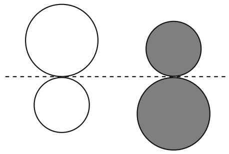


Fig. 8 A two-disk configuration that requires 4 translation moves

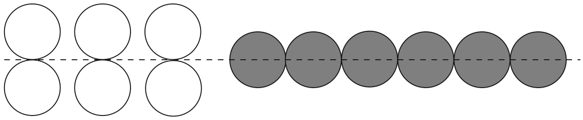


Fig. 9 A configuration of 6 congruent disks that requires 9 translation moves

being necessary (as previously argued for Fig. 6), now about $n/2 + n/4 + \cdots + n/2^k$ nontarget moves are necessary. The precise calculation for $m = 1$ yields the lower bound $2n - O(n^{\log_3 2}) = 2n - O(n^{0.631})$; see [8].

2.2 The Translation Model

This model, first studied by Abellanas et al. [2], is a constrained variant of the sliding model in which each move is a translation along a fixed direction; that is, the center of the moving disk traces a line segment. With some care, one can modify the universal algorithm mentioned in the introduction and find a suitable order in which disks can be moved "to infinity" and then moved "back" to target position via translations almost parallel to any given fixed direction using $2n$ translation moves [2].

That this bound is best possible for arbitrary radii disks can be easily seen in Fig. 8. Consider a pair of start and target configurations with two disks each. The two start disks and the two target positions are tangent to the same line. Note that the first move cannot be a target move. Assume that the larger disk moves first, and observe that its location must be above the horizontal line. If the second move is again a nontarget move, we have accounted for 4 moves already. Otherwise, no matter which disk moves to its target position, the other disk will require two more moves to reach its target. The situation when the smaller disk moves first is analogous. One can repeat this basic configuration with two disks, using different radii, to obtain configurations with an arbitrary large (even) number of disks, which require $2n$ translation moves.

Theorem 4 ([2]). *Given a pair of start and target configurations, each consisting of n disks of arbitrary radii, $2n$ translation moves always suffice for transforming the start configuration into the target configuration. On the other hand, there exist pairs of configurations that require $2n$ such moves.*

For congruent disks, the configuration shown in Fig. 9 (the first lower bound that was proposed) requires $3n/2$ moves, since from each pair of tangent disks, the first move must be a nontarget move [42]. A better lower bound, $\lfloor 8n/5 \rfloor$, due

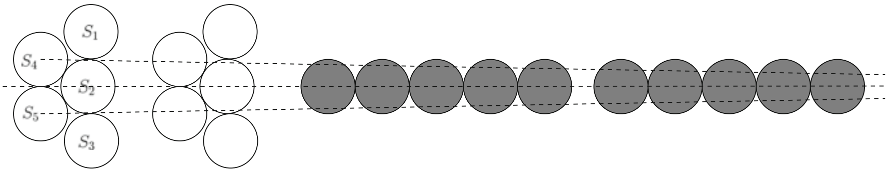


Fig. 10 Reconfiguration of a system of 10 congruent disks that needs 16 translation moves. Start disks are *white* and target disks are *shaded*

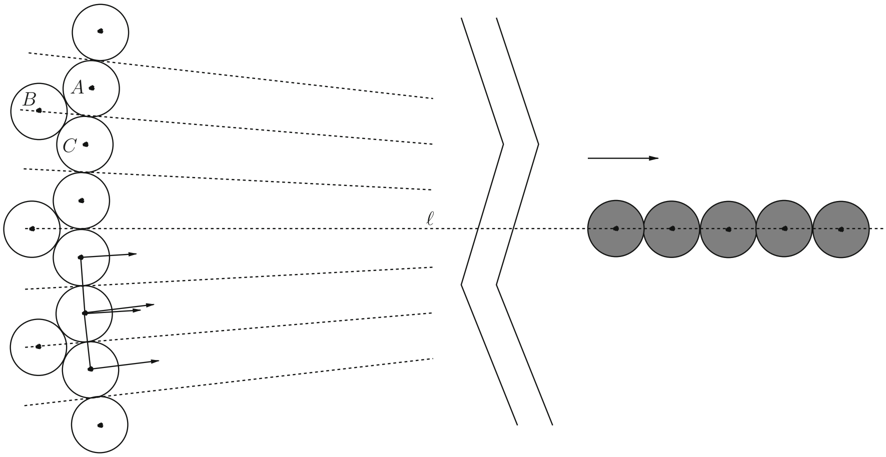


Fig. 11 Illustration of the lower-bound construction for translating congruent unlabeled disks, for $m = 3$, $n = 11$. The disks are *white* and their targets are *shaded*. Two consecutive partially overlapping parallel strips of width 2 are shown

to Abellanas et al. [2] is illustrated in Fig. 10. The construction is symmetric with respect to the middle horizontal line. Here we have groups of five disks each, where to "move" one group to some five target positions requires eight translation moves. In each group, the disks S_2, S_4, and S_5 are pairwise tangent, and S_1 and S_3 are each tangent to S_2; the tangency lines in the latter pairs are almost horizontal, converging to the middle horizontal line. Here are two different ways for moving one group, each requiring three nontarget moves: (a) S_1 and S_3 move out, S_2 moves to destination, S_4 moves out, S_5 moves to destinations followed by the rest. (b) S_4, S_5, and S_2 move out (to the left), S_1 and S_3 move to destinations followed by the rest.

The current best lower bound, $\lfloor 5n/3 \rfloor - 1$, is due to Dumitrescu and Jiang [24]. Let $n = 3m + 2$. The start and target configurations, each with n disks, are shown in Fig. 11. The n target positions are all on a horizontal line ℓ, with the disks at these positions forming a horizontal chain, $T_1, \ldots, T_n$, consecutive disks being tangent to each other. Let o denote the center of the median disk, $T_{\lfloor n/2 \rfloor}$. Let $r > 0$ be very large. The start disks are placed on two very slightly convex chains (two concentric circular arcs):

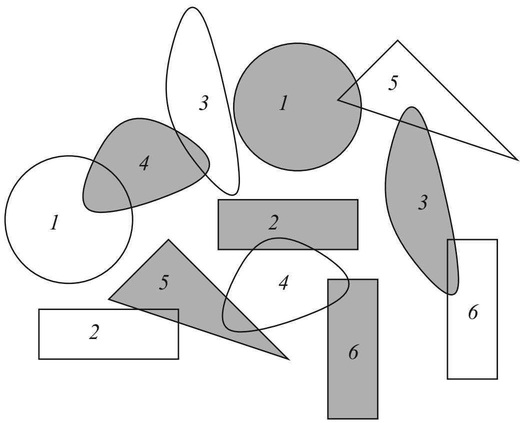


Fig. 12 Reconfiguration of convex bodies with translations. The start positions are *unshaded*; the target positions are *shaded*

- $2m+2$ disks in the first layer (chain). Their centers are $2m+2$ equidistant points on a circular arc of radius r centered at o.
- m disks in the second layer. Their centers are m equidistant points on a concentric circular arc of radius $r\cos\alpha+\sqrt{3}$. Each pair of consecutive points on the circle of radius r subtends an angle of 2α from the center of the circle (α is very small).

The parameters of the construction are chosen to satisfy $\sin\alpha = 1/r$ and $2n\sin n\alpha \le 2$. Set, for instance, $\alpha = 1/n^2$, which results in $r = \Theta(n^2)$. Alternatively, the configuration can be viewed as consisting of m groups of three disks each, plus two disks, one at the higher and one at the lower end of the chain along the circle of radius r. We summarize the bounds for translation of congruent disks in the following theorem.

Theorem 5 ([2, 24]). *Given a pair of start and target configurations, each consisting of n congruent disks,* $2n-1$ *translation moves always suffice for transforming the start configuration into the target configuration. On the other hand, there exist pairs of configurations that require* $\lfloor 5n/3 \rfloor - 1$ *such moves.*

Translating Convex Bodies. We briefly discuss the general problem of reconfiguration of convex bodies with translations. Refer to Fig. 12. When the convex bodies have different shapes, sizes, and orientations, we assume that the correspondence between the start positions $\{S_1,\ldots,S_n\}$ and the target positions $\{T_1,\ldots,T_n\}$ is given explicitly: T_i is a translated copy of S_i. In other words, we deal with the labeled variant of the problem. Theorem 6 can be easily extended to the unlabeled variant by first computing a valid correspondence by shape matching. The $2n$ upper bound for translating arbitrary disks can be extended to arbitrary convex bodies in the plane.

Theorem 6 ([24]). *For the reconfiguration with translations of n labeled disjoint convex bodies in the plane,* $2n$ *moves are always sufficient and sometimes necessary.*

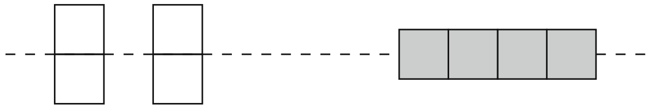


Fig. 13 A lower bound of $\lfloor 3n/2 \rfloor$ for translating axis-parallel unit squares. The start positions (grouped in pairs) are tangent to the x-axis, which intersects the target positions (*shaded*). Each of the target squares is symmetric about the x-axis

Translating Unlabeled Axis-Parallel Unit Squares. Throughout a translation move, the moving square remains axis-parallel; however, the move can be in any direction. The construction in Fig. 13 gives a lower bound of $\lfloor 3n/2 \rfloor$ and we have the following bounds.

Theorem 7 ([24]). *For the reconfiguration with translations of n unlabeled axis-parallel unit squares in the plane,* $2n-1$ *moves are always sufficient, and* $\lfloor 3n/2 \rfloor$ *moves are sometimes necessary.*

Recently García et al. estimated the number of moves necessary for the reconfiguration of n axis-parallel rectangles, where each move is a collision-free axis-parallel translation [30].

2.3 The Lifting Model

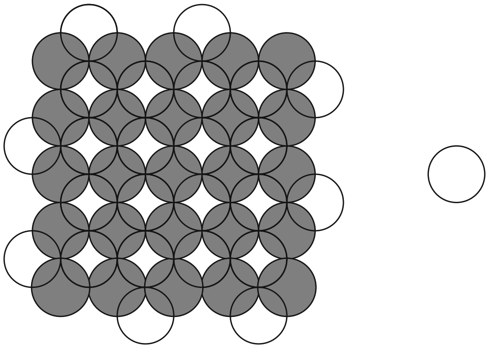
For systems of n congruent disks, one can estimate the number of moves that are always sufficient with higher accuracy. The following bounds were established by Dumitrescu and Bereg [7]:

Theorem 8 ([7]). *Given a pair of start and target configurations S and T, each with n congruent disks, one can move the disks from S to T using* $n+O(n^{2/3})$ *moves in the lifting model. The entire motion can be computed in* $O(n\log n)$ *time. On the other hand, for each n, there exist pairs of configurations that require* $n+\Omega(n^{1/2})$ *moves for this task.*

The lower-bound construction is illustrated in Fig. 14 for $n = 25$. Assume for simplicity that $n = m^2$, where m is odd. We place the disks of T onto a grid $X \times X$ of size $m \times m$, where $X = \{2,4,\ldots,2m\}$. We place the disks of S onto a grid of size $(m-1)\times(m-1)$ so that they overlap with the disks from T. The grid of target disks contains $4m-4$ disks on its boundary. We "block" them with $2m-2$ start disks in S by placing them so that each start disk overlaps with two boundary target disks as shown in the figure. We place the last start disk somewhere else, and we have accounted for $(m-1)^2+(2m-2)+1 = m^2$ start disks. As proved in [7], at least $n+\lfloor m/2 \rfloor$ moves are necessary for reconfiguration (it can be verified that this number of lifting moves suffices for this construction).

The proof of the upper bound of $n+O(n^{2/3})$ is technically more complicated. It builds a binary space partition of the plane into convex polygonal (bounded or

Fig. 14 A pair of start and target configurations, each with $n = 25$ congruent disks, which require 27 lifting moves. The start disks are *white* and the target disks are *shaded*



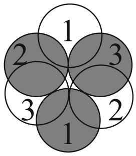


Fig. 15 A group of three disks (of distinct radii) that require five moves to reach their targets; part of the lower-bound construction for lifting disks of arbitrary radii. The disks are *white* and their targets are *shaded*

unbounded) regions satisfying certain properties. Once the partition is obtained, a *shifting* algorithm moves disks from some regions to fill the target positions in other regions; see [7] for details. Since the disks whose target position coincides with their start position can be safely ignored in the beginning, the upper bound yields an efficient algorithm that performs a number of moves close to the optimum (for large n).

For arbitrary radius disks, the following result is obtained in [7].

Theorem 9 ([7]). *Given a pair of start and target configurations S and T, each with n disks with arbitrary radii, $9n/5$ moves always suffice for transforming the start configuration into the target configuration. On the other hand, for each n, there exist pairs of configurations that require $\lfloor 5n/3 \rfloor$ moves for this task.*

The lower bound is very simple. We use disks of different radii (although the radii can be chosen very close to the same value if desired). Since all disks have distinct radii, one can think of the disks as being labeled. Consider the set of three disks, labeled 1, 2, and 3 in Fig. 15. The two start and target disks labeled i are congruent, for $i = 1, 2, 3$. To transform the start configuration into the target configuration takes at least two nontarget moves, thus five moves in total. By repeatedly using groups of three (with different radii), one gets a lower bound of $5n/3$ moves, when n is a multiple of three, and $\lfloor 5n/3 \rfloor$ in general.

We now explain the approach in [7] for the upper bound for n disks of arbitrary radii. Let $S = \{s_1, \ldots, s_n\}$ and $T = \{t_1, \ldots, t_n\}$ be the start and target configurations.

We assume that for each i, disk s_i is congruent to disk t_i; i.e., t_i is the target position of s_i. If the correspondence $s_i \to t_i$ is not given (but only the two sets of disks), it can be easily computed by sorting both S and T by radius.

In a directed graph $D = (V,E)$, let $d_v = d_v^+ + d_v^-$ denote the degree of vertex v, where d_v^+ is the out-degree of v and d_v^- is the in-degree of v. Let $\beta(D)$ be the maximum size of a subset V' of V such that $D[V']$, the subgraph induced by V', is acyclic. In [7] the following inequality is proved for any directed graph:

$$\beta(D) \geq \max\left(\sum_{v\in V} \frac{1}{d_v^+ + 1}, \sum_{v\in V} \frac{1}{d_v^- + 1}\right).$$

For a disk ω, let $\overset{\circ}{\omega}$ denote the interior of ω. Let S be a set of k pairwise disjoint red disks, and T be a set of l pairwise disjoint blue disks. Consider the bipartite red–blue disk intersection graph $G = (S,T,E)$, where $E = \{(s,t) \,:\, s \in S, t \in T, \overset{\circ}{s} \cap \overset{\circ}{t} \neq \emptyset\}$. Using the triangle inequality (among sides and diagonals in a convex quadrilateral), one can easily show that any red–blue disk intersection graph $G = (S,T,E)$ is planar, and consequently $|E| \leq 2(|S|+|T|) - 4 = 4n - 4$. We think of the start and target disks being labeled from 1 to n, so that the target of start disk i is target disk i. Consider the directed *blocking graph* $D = (S,F)$ on the set S of n start disks, where

$$F = \{(s_i, s_j) \,:\, i \neq j \text{ and } \overset{\circ}{s_i} \cap \overset{\circ}{t_j} \neq \emptyset\}.$$

If $(s_i, s_j) \in F$, we say that disk i *blocks* disk j. (Note that $s_i \cap t_i \neq \emptyset$ does not generate any edge in D.) Since if $(s_i, s_j) \in F$, then $(s_i, t_j) \in E$, we have $|F| \leq |E| \leq 4n - 4$. The algorithm first eliminates all the directed cycles from D using some nontarget moves, and then fills the remaining targets using only target moves. Let

$$d^+ = \frac{\sum_{v\in S} d_v^+}{n} = \frac{|F|}{n}$$

be the average out-degree in D. We have $d^+ \leq 4$, which further implies (by Jensen's inequality)

$$\beta(D) \geq \sum_{v\in S} \frac{1}{d_v^+ + 1} \geq \frac{n}{d^+ + 1} \geq \frac{n}{5}.$$

Let $S' \subset S$ be a set of disks of size at least $n/5$ and whose induced subgraph is acyclic in D. Move out far away the remaining set S'' of at most $4n/5$ disks, and note that the faraway disks do not block any of the disks in S'. Perform a topological sort on the acyclic graph $D[S']$ induced by S', and fill the targets of these disks in that order using only target moves. Then fill the targets with the faraway disks in any order. The number of moves is at most $n + 4n/5 = 9n/5$, as claimed. Figure 16 shows the bipartite intersection graph G and the directed blocking graph D for a small example, with the corresponding reconfiguration procedure explained above.

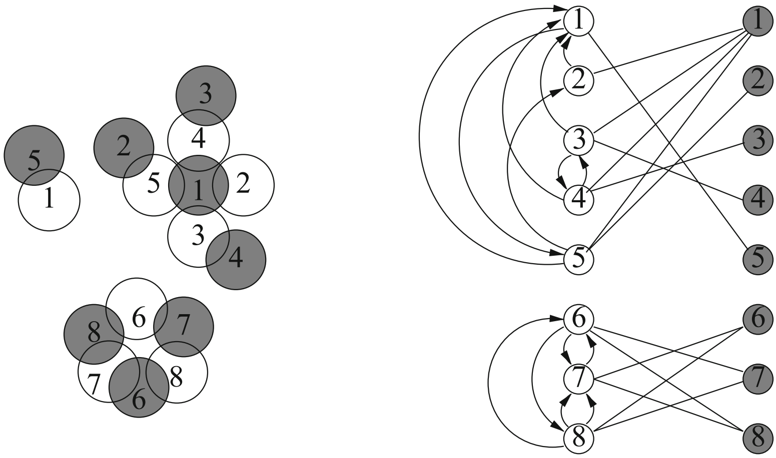


Fig. 16 The bipartite intersection graph G and the directed blocking graph D. Move out 4, 5, 7, 8; no cycles remain in D. Fill targets 3, 2, 1, 6, and then 4, 5, 7, 8. The start disks are *white* and the target disks are *shaded*

Similar to the case of congruent disks, the resulting algorithm performs a number of moves that is not more than a constant times the optimum. Indeed, as mentioned in the introduction, the disks whose target positions coincide with their start positions can be safely ignored. So each of the remaining start disk needs to be moved from its initial place. Without loss of generality, all disks need to move, so the number of moves in any solution is at least n (the case when some disks are ignored is only better). Since the algorithm above performs at most $9n/5$ lifting moves, it achieves a constant approximation ratio, $9/5$.

Computational Complexity of Optimal Reconfiguration. While, as discussed, reconfiguration with $2n$ or fewer moves is always possible in any of the three models (sliding, translation, and lifting), optimal reconfiguration (employing a minimum number of moves) is probably NP-hard in each of these models. This has been confirmed for translation and sliding by Dumitrescu and Jiang [24]: (a) Both the labeled and unlabeled versions of the disk reconfiguration problem with translations U-TRANS-RP and L-TRANS-RP are NP-hard even for congruent disks. (b) Both the labeled and unlabeled versions of the sliding disks reconfiguration problem in the plane U-SLIDE-RP and L-SLIDE-RP are NP-hard even for congruent disks.

2.4 Further Questions

We list a few open problems concerning the three models discussed:

1. Reduce the gap between the $16n/15 - o(n)$ lower bound and the $3n/2 + o(n)$ upper bound on the number of moves for sliding n congruent disks.

2. Consider the reconfiguration problem for congruent squares (with arbitrary orientation) in the sliding model. It can be checked that the $3n/2 + o(n)$ upper bound for congruent disks still holds in this case; however, the $16n/15 - o(n)$ lower bound based on stable disk packings cannot be used. Observe that the $n + \Omega(n^{1/2})$ lower bound for congruent disks in the lifting model (Fig. 14) can be realized with congruent (even axis-aligned) squares and therefore holds for congruent squares in the sliding model as well. Can one deduce better bounds for this variant?
3. Derive a $(2-\delta)n$ upper bound for translating congruent disks (where δ is a positive constant), or improve the (multiplicative constant in the) $\lfloor 5n/3 \rfloor - 1$ lower bound.
4. Consider the reconfiguration problem for congruent *labeled* disks in the sliding model. It is easy to see that the $\lfloor 5n/3 \rfloor$ lower bound holds, since the construction in Fig. 15 can be realized with congruent disks. Find a $(2-\delta)n$ upper bound (where δ is a positive constant), or improve the (multiplicative constant in the) $\lfloor 5n/3 \rfloor$ lower bound.
5. The type of construction in Fig. 13 has been used previously for disks to obtain the first lower bound of $\lfloor 3n/2 \rfloor$ for translating unit disks [42]. It is interesting to note that neither of the two subsequent improved constructions, $\lfloor 8n/5 \rfloor$ of Abellanas et al. [2], or ours $\lfloor 5n/3 \rfloor - 1$ in Theorem 5, seems to work for squares.
6. Reduce the gap between the $\lfloor 5n/3 \rfloor$ lower bound and the $9n/5$ upper bound on the number of moves for lifting n disks of arbitrary radii.
7. What is the computational complexity of the reconfiguration problem in the lifting model? Are U-LIFT-RP and L-LIFT-RP NP-hard for unit disks?

3 Modular and Reconfigurable Systems

A number of issues related to motion planning and analysis of rectangular metamorphic robotic systems are addressed in [26]. A distributed algorithm for reconfiguration that applies to a relatively large subclass of configurations, called horizontally convex configurations, is presented. Several fundamental questions in the analysis of metamorphic systems have been also addressed. In particular, the following two questions have been shown to be decidable: (a) whether a given set of motion rules maintains connectivity; (b) whether a goal configuration is reachable from a given initial configuration (at specified locations).

For illustration, we present the *rectangular model* of metamorphic systems introduced in [25–27]. Consider a plane that is partitioned into an integer grid of square cells indexed by their center coordinates in the underlying x–y coordinate system. This partition of the plane is only a useful abstraction, as the module-size determines the grid size in practice. At any time, each cell may be empty or occupied by a module. The *reconfiguration* of a metamorphic system consisting of n modules is a sequence of configurations of the modules in the grid at discrete time steps $t = 0, 1, 2, \ldots$; see below. Let V_t be the configuration of the modules at time t, where

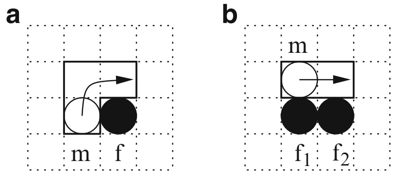


Fig. 17 Moves in the rectangular model: (**a**) clockwise *NE* rotation and (**b**) sliding in the *E* direction. Fixed modules are *shaded*. The cells in which the moves take place are outlined in the figure

we often identify V_t with the set of cells occupied by the modules or with the set of their centers. A useful feature we insist on is maintaining connectivity; i.e., for each t, the graph $G_t = (V_t, E_t)$ must be connected, where for any t, E_t is the set of edges connecting pairs of cells in V_t that are side-adjacent. The following two generic motion rules (Fig. 17) define the rectangular model [25–27]. These rules can be applied in all axis-parallel orientations, in fact generating 16 rules, eight for rotation and eight for sliding. A somewhat similar model is presented in [13].

- *Rotation*: A module m side-adjacent to a stationary module f rotates through an angle of 90° around f either clockwise or counterclockwise. Figure 17a shows a clockwise *NE* rotation. For rotation to take place, both the target cell and the cell at the corresponding corner of f that m passes through (*NW* in the given example) have to be empty.
- *Sliding*: Let f_1 and f_2 be stationary cells that are side-adjacent. A module m that is side-adjacent to f_1 and adjacent to f_2 slides along the sides of f_1 and f_2 into the cell that is adjacent to f_1 and side-adjacent to f_2. Figure 17b shows a sliding move in the *E* direction. For sliding to take place, the target cell has to be empty.

The system may execute moves sequentially, when only one module moves at each discrete time step, or concurrently (when more modules can move at each discrete time step). Concurrent execution has the advantage to speed up the reconfiguration process. If concurrent moves are allowed, additional conditions have to be imposed, as in [26, 27]. In order to ensure motion precision, each move is guided by one or two modules that are stationary during the same step. The following result of Dumitrescu and Pach settles a conjecture formulated in [26].

Theorem 10 ([25]). *The set of motion rules of the rectangular model guarantees the feasibility of motion planning for any pair of connected configurations V and V' having the same number of modules. That is, following the above rules, V and V' can always be transformed into each other so that all intermediate configurations are connected.*

The algorithm in [25] is far from being intuitive or straightforward. The main difficulties that have to be overcome are dealing with holes and avoiding certain deadlock situations during reconfiguration. The proof of correctness of the algorithm and the analysis of the number of moves (cubic in the number of modules,

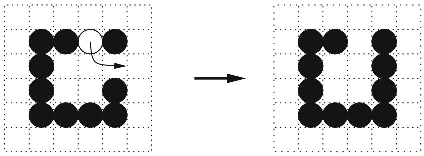


Fig. 18 A rotation move that temporarily disconnects the configuration

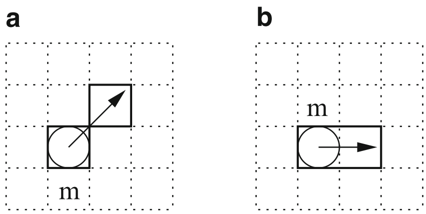


Fig. 19 Moves in the weak rectangular model: (**a**) *NE* diagonal move and (**b**) side move in the *E* direction. The cells in which the moves take place are outlined in the figure

for sequential execution) are also quite involved. It is easy to construct examples so that neither sliding nor rotation alone can reconfigure them to straight chains; here a *straight chain* is a set of modules that form a straight line chain in the grid. However, conform with Theorem 10, the motion rules of the rectangular model (rotation and sliding, Fig. 17) are sufficient to guarantee reachability while maintaining the system connected at each discrete time step. This had been proved earlier only for the special class of horizontally convex systems [26].

A somewhat different model can be obtained if, instead of the connectedness requirement at each time step, one imposes the *single backbone* condition [26]: Each move of a module (sliding or rotation) is along the single connected backbone formed by the other modules. If concurrent moves are allowed, additional conditions have to be imposed, as in [26]. A subtle difference exists between requiring the configuration to be connected at each discrete time step and requiring the existence of a connected backbone along which a module slides or rotates. A one step motion that does not satisfy the single backbone condition appears in Fig. 18: The initial connected configuration temporarily disconnects during the move and reconnects at the end of it. The algorithm from [25] has the nice property that the single backbone condition is satisfied during the whole procedure.

It is worth briefly discussing another rectangular model for which the result in Theorem 10 holds. The following two generic motion rules (Fig. 19) define the *weak rectangular model*. These rules are again applicable in all axis-parallel orientations, in fact generating eight rules, four diagonal moves and four side (axis-parallel) moves. The only requirement is that the configurations must remain connected at each discrete time step.

- *Diagonal move*: A module m moves diagonally to an empty cell corner adjacent to $cell(m)$.
- *Side move*: A module m moves to an empty cell side adjacent to cell(m).

The same result from Theorem 10 holds for this second model [25]; however, its proof and corresponding reconfiguration algorithm are much simpler.

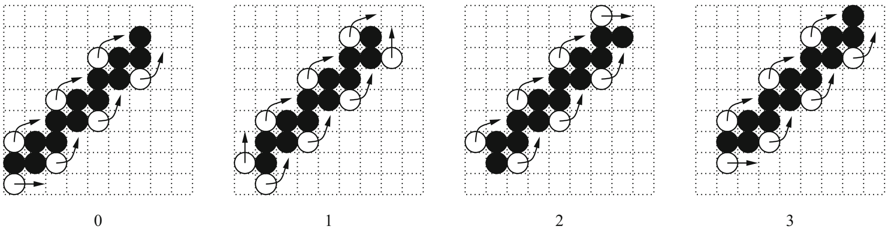


Fig. 20 Formation of 20 modules moving diagonally at a speed of $\frac{1}{3}$ (diagonal formation)

Theorem 11 ([25]). *The set of motion rules of the weak rectangular model guarantees the feasibility of motion planning for any pair of connected configurations having the same number of modules.*

It is worth mentioning that reconfigurations in the rectangular model can be also viewed in the broader context of transformations of binary images. Bose et al. recently studied several such transformations while insisting on maintaining the connectivity of both the foreground and the background of the images in each step [10].

A different variant of interrobot reconfiguration is useful in applications for which there is no clear preference between the use of a single large robot versus a group of smaller ones [15]. This leads to the merging of individual smaller robots into a larger one or the splitting of a large robot into smaller ones. For example, in a surveillance or rescue mission, a large robot is required to travel to a designated location in a short time. Then the robot may create a group of small robots that are to explore concurrently a large area. Once the task is complete, the robots might merge back into the large robot that carried them.

As mentioned in [22], there is considerable research interest in the task of having one autonomous vehicle follow another, and in general in studying robots moving in formation. Dumitrescu et al. [27] examined the problem of dynamic self-reconfiguration of a modular robotic system to a formation aimed at reaching a specified target position as quickly as possible. A number of fast formations for both rectangular and hexagonal systems are presented, achieving a constant-ratio guarantee on the time to reach a given target in the asymptotic sense. For example, a snake-like formation (with $n \geq 4$ modules, n even) can move at a speed of $\frac{1}{3}$ in the rectangular model. In Fig. 20 the formation at time 0 reappears at time 3 translated diagonally by one unit. Thus, by repeatedly going through these configurations, the modules can move in the *NE* direction at a speed of $\frac{1}{3}$.

We conclude this section with some remaining questions on modular and reconfigurable systems related to the results presented. Preliminary results of Abel and Kominers [1] suggest that the first two questions below have a positive answer. This, however, remains to be confirmed.

1. The reconfiguration algorithm in the rectangular model makes at most $2n^3$ moves in the worst case [26]. On the other hand, the reconfiguration of a vertical chain

into a horizontal chain requires only $\Theta(n^2)$ moves, and it is believed that no other pair of configurations requires more. A quadratic upper bound on the number of moves has been proved in the weak rectangular model [25], but it remains open in the first model.
2. Study whether the analogous rules of rotation and sliding in three dimensions permit the feasibility of motion planning for any pair of connected configurations having the same number of modules.
3. If concurrent execution is permitted (under appropriate assumptions), what is the smallest number of concurrent steps that suffices for the reconfiguration of any pair of connected configurations V and V' with n modules—in the rectangular model and in the weak variant? Is it $O(n)$? (This has been shown for a special class of horizontally convex systems [26] and, e.g., in a related 3D rectangular model studied in [4].)

4 Reconfigurations in Graphs and Grids

In certain applications, objects are indistinguishable, and the chips are therefore unlabeled; for instance, a modular robotic system consists of a number of identical modules (robots), each having identical capabilities [25–27]. In other applications the chips may be labeled. The variant with unlabeled chips is easier and always feasible, while the variant with labeled chips may be infeasible: A classical example is the 15-puzzle on a 4×4 grid—introduced by Sam Loyd in 1878—which admits a solution if and only if the start permutation is an even permutation [33, 47]. Most of the work done so far concerns labeled versions of the reconfiguration problem, and here we give only a short account.

For the generalization of the 15-puzzle on an arbitrary graph (with $k = v - 1$ labeled chips in a connected graph on v vertices), Wilson [51] gave an efficiently checkable characterization of the solvable instances of the problem. Kornhauser et al. have extended his result to any $k \leq v - 1$ and provided bounds on the number of moves for solving any solvable instance [35]. Ratner and Warmuth showed that finding a solution with a minimum number of moves for the $(N \times N)$-extension of the 15-puzzle is NP-hard [43], so the reconfiguration problem in graphs with labeled chips is NP-hard. Auletta et al. gave a linear-time algorithm for the *pebble motion on a tree* problem [5]. This problem is the labeled variant of the same reconfiguration problem studied in [16]; however, each move is along one edge only. Papadimitriou et al. studied a problem of *motion planning on a graph* in which there is a mobile robot at one of the vertices, say s, that has to reach a designated vertex t using the smallest number of moves, in the presence of obstacles (pebbles) at some of the other vertices [41]. Robot and obstacle moves are done along edges, and obstacles have no destination assigned and may end up in any vertex of the graph. The problem has been shown to be NP-complete even for planar graphs, and a polynomial-time approximation algorithm with ratio $O(\sqrt{n})$ was given in [41].

The following results are proved in [16] for the "chips in graph" reconfiguration problem described in part (3) of Sect. 1. Recall that in one move a chip can follow a path in the graph and end up at another vertex, provided the path (including the end vertex) is free of other chips.

(a) The reconfiguration problem in graphs with unlabeled chips U-GRAPH-RP is NP-hard, and even APX-hard.
(b) The reconfiguration problem in graphs with labeled chips L-GRAPH-RP is APX-hard.
(c) For the infinite planar rectangular grid, both the labeled and unlabeled variants L-GRID-RP and U-GRID-RP are NP-hard.
(d) There is a ratio-3 approximation algorithm for the unlabeled version in graphs U-GRAPH-RP. Thereby one gets a ratio-3 approximation algorithm for the unlabeled version U-GRID-RP in the (infinite) rectangular grid.
(e) It can be shown that n moves are always enough (and sometimes necessary) for the reconfiguration of n unlabeled chips in graphs. For the case of trees, a linear-time algorithm that performs an optimal (minimum) number of moves is presented.
(f) It is shown that $7n/4$ moves are always enough, and $3n/2$ are sometimes necessary, for the reconfiguration of n labeled chips in the infinite planar rectangular grid (L-GRID-RP).

Next, we give some details showing that in the infinite grid, n moves always suffice for the reconfiguration of n unlabeled chips, and of course it is easy to construct tight examples. The result holds in a more general graph setting [item (5) in the above list]: Let G be a connected graph, and let V and V' two n-element subsets of its vertex set $V(G)$. Imagine that we place a chip at each element of V and we want to move them into the positions of V' (V and V' may have common elements). A move is defined as shifting a chip from v_1 to v_2 [$v_1, v_2 \in V(G)$] along a path in G so that no intermediate vertices are occupied.

Theorem 12 ([16]). *In G one can get from any n-element initial configuration V to any n-element final configuration V' using at most n moves, so that no chip moves twice.*

It suffices to prove the theorem for trees, and we'd like to include the short proof here. We argue by induction on the number of chips. Take the smallest tree T containing V and V', and consider an arbitrary leaf l of T. Assume first that the leaf l belongs to V: say $l = v$. If v also belongs to V', the result trivially follows by induction, so assume that this is not the case. Choose a path P in T, connecting v to an element v' of V' such that no internal point of P belongs to V'. Apply the induction hypothesis to $V \setminus \{v\}$ and $V' \setminus \{v'\}$ to obtain a sequence of at most $n-1$ moves, and add a final (unobstructed) move from v to v'.

The remaining case when the leaf l belongs to V' is symmetric: Say $l = v'$; choose a path P in T, connecting v' to an element v of V such that no internal point of P belongs to V. First move v to v' and append the sequence of at most $n-1$ moves obtained from the induction hypothesis applied to $V \setminus \{v\}$ and $V' \setminus \{v'\}$. This completes the proof.

Theorem 12 implies that in the infinite rectangular grid, we can get from any starting position to any ending position of the same size n in at most n moves. It is interesting to compare this to the problem of sliding congruent unlabeled disks in the plane, where one can come up with cage-like constructions that require about $\frac{16n}{15}$ moves [8], as discussed in Sect. 2.1. We conclude this section with two remaining questions on reconfigurations in graphs and grids:

1. Can the ratio-3 approximation algorithm for the unlabeled version in graphs U-GRAPH-RP be improved? Is there an approximation algorithm with a better ratio for the infinite planar rectangular grid?
2. Is it possible to close or reduce the gap between the $3n/2$ lower bound and the $7n/4$ upper bound on the number of moves for the reconfiguration of n labeled chips in the infinite planar rectangular grid?

5 Conclusion

The different reconfiguration models discussed in this survey have raised new interesting mathematical questions and revealed surprising connections with other older ones. For instance, the key ideas in the reconfiguration algorithm in [25] were derived from the properties of a system of *maximal cycles*, similar to those of the block decomposition of graphs [21]. The lower-bound configuration with unit disks for the sliding model in [8] uses "rigidity" considerations and properties of *stable* packings of disks studied a long time ago by Böröczky [9]; in particular, he showed the existence of stable systems of unit disks with arbitrarily small density. A suitable modification of his construction yields our lower bound.

The study of the lifting model offered other interesting connections: The algorithm for unit disks given in [7] is intimately related to the notion of a *center point* of a finite point set and to the following property derived from it: Given two sets each with n pairwise disjoint unit disks, there exists a binary space partition of the plane into polygonal regions each containing roughly the same small number ($\approx n^{2/3}$) of disks and such that the total number of disks intersecting the boundaries of the regions is small ($\approx n^{2/3}$). The reconfiguration algorithm for disks of arbitrary radius relies on a new lower bound on the maximum order of induced *acyclic* subgraphs of a directed graph [7], analogous to the bound on the *independence number* of an undirected graph given by Turán's theorem [3]. Moreover, we have used the crucial fact that a bipartite disk intersection graph (drawn as a geometric graph on the set of disk centers) is planar, to obtain a linear upper bound on its number of edges. Finally, the ratio-3 approximation algorithm for the unlabeled version in graphs is obtained by applying the *local ratio method* of Bar-Yehuda [6] to a graph H constructed from the given graph G.

Regarding the various models of reconfiguration for systems of objects in the plane, we have presented estimates on the number of moves that are necessary in the worst case. From the practical viewpoint, one would like to convert these estimates into approximation algorithms with a good ratio guarantee. As shown for the lifting model, the upper-bound estimates on the number of moves give good approximation algorithms for large values of n. However, further work is needed in this direction for the sliding model and the translation model in particular.

Note. An earlier survey [23] on this topic came out with many misprints and errors that were introduced by the publisher. This prompted the author to prepare the current updated version.

Acknowledgements The author is grateful to his collaborators, S. Bereg, G. Călinescu, M. Jiang, J. Pach, I. Suzuki, and M. Yamashita, for allowing him to borrow some material from their respective joint works.

References

1. Z. Abel, S.D. Kominers, Pushing hypercubes around. Preprint (2008). `arXiv: 0802.3414v2`
2. M. Abellanas, S. Bereg, F. Hurtado, A.G. Olaverri, D. Rappaport, J. Tejel, Moving coins. Comput. Geom.: Theor. Appl. **34**(1), 35–48 (2006)
3. N. Alon, J. Spencer, *The Probabilistic Method*, 2nd edn. (Wiley, New York, 2000)
4. G. Aloupis, S. Collette, M. Damian, E.D. Demaine, R.Y. Flatland, S. Langerman, J. O'Rourke, S. Ramaswami, V. Sacristán Adinolfi, S. Wuhrer, Linear reconfiguration of cube-style modular robots. Comput. Geom.: Theor. Appl. **42**(6–7), 652–663 (2009)
5. V. Auletta, A. Monti, M. Parente, P. Persiano, A linear-time algorithm for the feasibility of pebble motion in trees. Algorithmica **23**, 223–245 (1999)
6. R. Bar-Yehuda, One for the price of two: a unified approach for approximating covering problems. Algorithmica **27**, 131–144 (2000)
7. S. Bereg, A. Dumitrescu, The lifting model for reconfiguration. Discr. Comput. Geom. **35**(4), 653–669 (2006)
8. S. Bereg, A. Dumitrescu, J. Pach, Sliding disks in the plane. Int. J. Comput. Geom. Appl. **15**(8), 373–387 (2008)
9. K. Böröczky, Über stabile Kreis- und Kugelsysteme. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. **7**, 79–82 (1964) [in German]
10. P. Bose, V. Dujmović, F. Hurtado, P. Morin, Connectivity-preserving transformations of binary images. Comput. Vis. Image Understand. **113**, 1027–1038 (2009)
11. P. Braß, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, New York, 2005)
12. N.G. de Bruijn, Aufgaben 17 and 18. Nieuw Archief voor Wiskunde **2**, 67 (1954) (in Dutch)
13. Z. Butler, K. Kotay, D. Rus, K. Tomita, Generic decentralized control for a class of self-reconfigurable robots, in *Proceedings of the 2002 IEEE International Conference on Robotics and Automation* (ICRA'02), Washington, DC, May 2002, pp. 809–816
14. A. Casal, M. Yim, Self-reconfiguration planning for a class of modular robots, in *Proceedings of the Symposium on Intelligent Systems and Advanced Manufacturing (SPIE'99)*, 1999, pp. 246–256
15. A. Castano, P. Will, A polymorphic robot team, in *Robot Teams: From diversity to Polymorphism*, ed. by T. Balch, L.E. Parker (A K Peters, 2002), pp. 139–160

16. G. Călinescu, A. Dumitrescu, J. Pach, Reconfigurations in graphs and grids. SIAM J. Discr. Math. **22**(1), 124–138 (2008)
17. B. Chazelle, T.A. Ottmann, E. Soisalon-Soininen, D. Wood, The complexity and decidability of separation, in *Proceedings of the 11th International Colloquium on Automata, Languages and Programming (ICALP '84)*. LNCS, vol. 172 (Springer-Verlag, New York, 1984), pp. 119–127
18. G. Chirikjian, A. Pamecha, I. Ebert-Uphoff, Evaluating efficiency of self-reconfiguration in a class of modular robots. J. Robot. Syst. **13**(5), 317–338 (1996)
19. H.T. Croft, K.J. Falconer, R.K. Guy, *Unsolved Problems in Geometry* (Springer-Verlag, New York, 1991)
20. E.D. Demaine, R.A. Hearn, Games, puzzles and computation. A K Peters: I-IX, 1–237 (2009)
21. R. Diestel, *Graph Theory, 2nd edn., Graduate Texts in Mathematics*, vol. 173 (Springer-Verlag, New York, 2000)
22. G. Dudek, M. Jenkin, E. Milios, A taxonomy of multirobot systems, in *Robot Teams: From diversity to Polymorphism*, ed. by T. Balch, L.E. Parker (A K Peters, 2002), pp. 3–22
23. A. Dumitrescu, Motion planning and reconfiguration for systems of multiple objects, in *Mobile Robots: Perception & Navigation*, ed. by S. Kolski. Advanced Robotic Systems, Kirchengasse 43/3, A-1070 Vienna, Austria, EU, 523–542 (2007)
24. A. Dumitrescu, M. Jiang, On reconfiguration of disks in the plane and other related problems, in *Proceedings of the 23rd International Symposium on Algorithms and Data Structures (WADS'09)*, Banff, Alberta, Canada, August 2009. LNCS, vol. 5664 (Springer, New York, 2009), pp. 254–265
25. A. Dumitrescu, J. Pach, Pushing squares around. Graphs Comb. **22**(1), 37–50 (2006)
26. A. Dumitrescu, I. Suzuki, M. Yamashita, Motion planning for metamorphic systems: feasibility, decidability and distributed reconfiguration. IEEE Trans. Robot. Autom. **20**(3), 409–418 (2004)
27. A. Dumitrescu, I. Suzuki, M. Yamashita, Formations for fast locomotion of metamorphic robotic systems. Int. J. Robot. Res. **23**(6), 583–593 (2004)
28. M. Erdmann, T. Lozano-Pérez, On multiple moving objects. Algorithmica **2**(1–4), 477–521 (1987)
29. L. Fejes Tóth, A. Heppes, Über stabile Körpersysteme. Compos. Math. **15**, 119–126 (1963) [in German]
30. A. García, C. Huemer, F. Hurtado, J. Tejel, Moving rectangles, in *Proceedings of the VII Jornadas de Matemática Discreta y Algorítmica*, Santander, 2010
31. L. Guibas, F.F. Yao, On translating a set of rectangles, in *Computational Geometry*, ed. by F. Preparata. Advances in Computing Research, vol. 1 (JAI Press, London, 1983), pp. 61–67
32. Y. Hwuang, N. Ahuja, Gross motion planning—a survey. ACM Comput. Surv. **25**(3), 219–291 (1992)
33. W.W. Johnson, Notes on the 15 puzzle. I. Am. J. Math. **2**, 397–399 (1879)
34. K. Kedem, M. Sharir, An efficient motion-planning algorithm for a convex polygonal object. Discr. Comput. Geom. **5**, 43–75 (1990)
35. D. Kornhauser, G. Miller, P. Spirakis, Coordinating pebble motion on graphs, the diameter of permutation groups, and applications, in *Proceedings of the 25th Symposium on Foundations of Computer Science (FOCS'84)*, Singer Island, FL, pp. 241–250 (1984)
36. L. Moser, Problem 66–11. SIAM Rev. **8**, 381 (1966)
37. S. Murata, H. Kurokawa, S. Kokaji, Self-assembling machine, in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA'94)*, San Diego, CA, USA, pp. 441–448 (1994)
38. A. Nguyen, L.J. Guibas, M. Yim, Controlled module density helps reconfiguration planning, in *Proceedings of IEEE International Workshop on Algorithmic Foundations of Robotics (WAFR'00)*, Dartmouth College Hanover, NH (2000)
39. J. Pach, M. Sharir, *Combinatorial Geometry and Its Algorithmic Applications—The Alcalá Lectures* (American Mathematical Society, Providence, RI, 2009)
40. A. Pamecha, I. Ebert-Uphoff, G. Chirikjian, Useful metrics for modular robot motion planning. IEEE Trans. Robot. Autom. **13**(4), 531–545 (1997)

41. C. Papadimitriou, P. Raghavan, M. Sudan, H. Tamaki, Motion planning on a graph, in *Proceedings of the 35th Symposium on Foundations of Computer Science (FOCS'94)*, pp. 511–520
42. D. Rappaport, *Communication at the 16th Canadian Conference on Computational Geometry (CCCG'04)*, August 2004
43. D. Ratner, M. Warmuth, Finding a shortest solution for the $(N \times N)$-extension of the 15-puzzle is intractable. J. Symb. Comput. **10**, 111–137 (1990)
44. J. O'Rourke, *Computational Geometry in C*, 2nd edn. (Cambridge University Press, Cambridge, 1998)
45. D. Rus, M. Vona, Crystalline robots: self-reconfiguration with compressible unit modules. Autonomous Robots **10**, 107–124 (2001)
46. J.T. Schwartz, M. Sharir, On the "piano mover's" problem, I. Commun. Pure Appl. Math. **36**, 345–398 (1983); II. Adv. Appl. Math. **4**, 298–351 (1983); III. Int. J. Robot. Res. **2**, 46–75 (1983); V. Commun. Pure Appl. Math. **37**, 815–848 (1984)
47. W.E. Story, Notes on the 15 puzzle. II. Am. J. Math. **2**, 399–404 (1879)
48. J.E. Walter, J.L. Welch, N.M. Amato, Distributed reconfiguration of metamorphic robot chains, in *Proceedings of the 19th ACM Symposium on Principles of Distributed Computing (PODC'00)*, July 2000, Portland, Oregon, pp. 171–180
49. J.E. Walter, J.L. Welch, N.M. Amato, Reconfiguration of hexagonal metamorphic robots in two dimensions, in *Sensor Fusion and Decentralized Control in Robotic Systems III*, ed. by G.T. McKee, P.S. Schenker. Proceedings of the Society of Photo-Optical Instrumentation Engineers, vol. 4196, 2000, pp. 441–453
50. X. Wei, W. Shi-gang, J. Huai-wei, Z. Zhi-zhou, H. Lin, Strategies and methods of constructing self-organizing metamorphic robots, in *Mobile Robots: New Research*, ed. by J.X. Liu (Nova Science, 2005), pp. 1–38
51. R.M. Wilson, Graph puzzles, homotopy, and the alternating group. J. Comb. Theor. Ser. B **16**, 86–96 (1974)
52. M. Yim, Y. Zhang, J. Lamping, E. Mao, Distributed control for 3D metamorphosis. Autonomous Robots **10**, 41–56 (2001)
53. E. Yoshida, S. Murata, A. Kamimura, K. Tomita, H. Kurokawa, S. Kokaji, A motion planning method for a self-reconfigurable modular robot, in *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Springer, Berlin, Germany (2001)

Rectangle and Square Representations of Planar Graphs

Stefan Felsner

Abstract In the first part of this survey, we consider planar graphs that can be represented by a dissections of a rectangle into rectangles. In rectangular drawings, the corners of the rectangles represent the vertices. The graph obtained by taking the rectangles as vertices and contacts as edges is the rectangular dual. In visibility graphs and segment contact graphs, the vertices correspond to horizontal or to horizontal and vertical segments of the dissection. Special orientations of graphs turn out to be helpful when dealing with characterization and representation questions. Therefore, we look at orientations with prescribed degrees, bipolar orientations, separating decompositions, and transversal structures.

In the second part, we ask for representations by a dissections of a rectangle into squares. We review results by Brooks et al. [The dissection of rectangles into squares. Duke Math J 7:312–340 (1940)], Kenyon [Tilings and discrete Dirichlet problems. Isr J Math 105:61–84 (1998)], and Schramm [Square tilings with prescribed combinatorics. Isr J Math 84:97–118 (1993)] and discuss a technique of computing squarings via solutions of systems of linear equations.

Mathematics Subject Classification (2010): 05C10, 05C62, 52C15.

1 Introduction

Questions around representations of graphs by geometric objects are intensively studied. Motivation comes from practical applications and the fascinating exchange between geometry and graph theory and other mathematical areas.

S. Felsner (✉)
Institut für Mathematik, Technische Universität Berlin, Berlin, Germany
e-mail: felsner@math.tu-berlin.de

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_12,

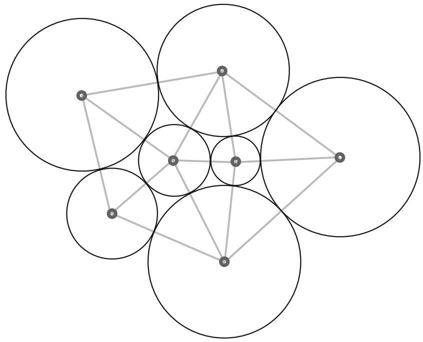


Fig. 1 A circle contact representation of a planar graph

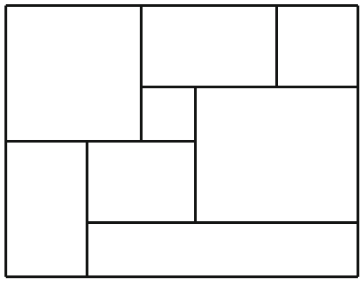


Fig. 2 A rectangular dissection

One of the nicest results about representations of graphs by geometric objects is Koebe's "coin graph theorem" [6, 35, 48]. It asserts that every planar graph can be represented by a set of disjoint discs, one for each vertex, such that two discs touch exactly if there is an edge between the corresponding vertices. Such a representation is called a *circle contact representation*. Figure 1 shows an example.

In the 1980s, Thurston observed a connection between circle packings and the Riemann mapping theorem. From there the theory has developed into a discrete analog of complex analysis; Stephenson's book [51] gives a comprehensive introduction.

In this chapter, we focus on representations of planar graphs based on rectangles. We look at rectangular dissections as shown in Fig. 2 and graphs that can be derived from it.

Suppose that $\phi : R \to G_\phi(R)$ corresponds to a specific mapping that associates a graph $G_\phi(R)$ with a rectangular dissection R and that $G_\phi(R)$ belongs to a class $\mathcal{G}_\phi$ of graphs. Then we can ask whether a given graph G from class $\mathcal{G}_\phi$ is *representable*, i.e., whether G is in the image of ϕ. The representability question can be treated as a characterization problem or as an algorithmic problem. If G is representable, we can also ask for a *representation*, i.e., for a dissection R such that $G = G_\phi(R)$. In this survey we consider several graphs associated to a dissection and the corresponding representability and representation problems.

In Sect. 2, we look at rectangular drawings and rectangular duals. Taking the corners of a rectangular dissection as vertices and the connecting line segments as edges yields a rectangular drawing. In Sect. 2.1, we review the theory of rectangular drawings, and in Sect. 2.3, we present an algorithm to decide whether a graph admits such a drawing and, if so, we generate it. The algorithm is based on an orientation of the angular graph with prescribed out-degrees. In some variants of the problem, this approach yields the fastest-known algorithm. In Sect. 2.2, we consider rectangular duals; in this model the vertices of the planar graph are represented by rectangles and edges by contacts.

In Sect. 3, we take the horizontal and vertical segments or just the horizontal segments of a rectangular dissection as vertices. Based on horizontal and vertical segments, we define the segment contact graph of a dissection and then proceed to consider more general segment contact graphs. With Theorem 3.2, we prove an unpublished characterization of segment 2-contact graphs due to Thomassen. In Sect. 3.1.1, segment contact graphs of rectangular dissections are shown to be closely related to separating decompositions. In Sect. 3.2, we consider the visibility graph of a dissection. This leads to the study of bipolar orientations. At the end of the section, in Sect. 3.3, we look at the relation between bipolar orientations and separating decompositions.

In the second part of the survey, we focus on square dissections, i.e., rectangular dissections where all rectangles are squares. Section 4 deals with the square analogs of visibility and segment contact graphs. We begin in Sect. 4.1 with the classical connection between squarings and electricity. In Sect. 4.2, we study a system of linear equations obtained from a separating decomposition and show that a solution yields a squaring. Kenyon [30] developed a more general theory relating trapezoidal dissections and Markov chains; it is the subject of Sect. 4.3.

Section 5 is based on Schramm [49]. The result is a characterization of graphs admitting a square dual. Finally, in Sect. 5.1, we relate square duals and transversal structures and propose an alternative method for computing square duals. The method is simple but comes with the drawback that its correctness still depends on a conjecture.

2 Rectangular Drawing and Rectangular Duals

2.1 Rectangular Drawing

Think of R as a union of interiorly disjoint rectangles. The union of the boundaries of the rectangles is the *skeleton* $\mathsf{skel}(R)$ of the dissection R. Let $C(R)$ be the set of corners of the rectangles of R. The skeleton of R can be viewed as a graph $G_{\mathsf{skel}}(R)$. The vertices of $G_{\mathsf{skel}}(R)$ are the points in $C(R)$, and the edges of $G_{\mathsf{skel}}(R)$ are the connecting line segments. More formally, the edges correspond to the connected components of $\mathsf{skel}(R) \setminus C(R)$. The *skeleton graph* $G_{\mathsf{skel}}(R)$ has four vertices of degree 2 incident to the outer face. All the other vertices are of degree 3 or 4. The edges are drawn as horizontal or vertical line segments.

If a graph G is represented by R, i.e., $G = G_{\mathsf{skel}}(R)$, then we call the representation a *rectangular drawing* of G. A characterization of graphs with $\Delta \leq 3$ that admit a rectangular drawing was obtained by Thomassen [54]. Thomassen's result is based on an earlier result of Ungar [59], who gave a characterization in the model where the corners of the outer face are regarded to be bends in an edge instead of vertices of degree 2. Ungar's characterization is dual to Theorem 2.2.

Algorithms for the construction of rectangular drawings have been considered by various authors. One of the key results is the following.

Theorem 2.1. *Let G be a plane graph with four distinguished corner vertices of degree 2 and vertices of degrees at most 3 otherwise. There is an algorithm that decides whether G is a skeleton graph and if so computes a rectangular drawing in linear time.*

A survey of algorithms and many references can be found in Chapter 6 of the book of Nishiseki and Rhaman [43]. In Sect. 2.3, we present an approach to rectangular drawings and a proof of Theorem 2.1 that is not covered there. This leads to improvements in the running times of some variants of the problem.

In the graph drawing literature, rectangular drawings have been extended and generalized in various directions.

- Edges are allowed to bend but remain constrained to the orthogonal drawing mode, i.e., are composed of horizontal and vertical segments. A highlight of the theory is the application of min-cost-flow algorithms for bend minimization pioneered by Tamassia [53].
- To overcome the degree restriction, some authors allow that in the drawing vertices are represented by boxes. With boxes and bends, every planar graph can be represented. If bends are forbidden, the problem can be reduced to finding a rectangular drawing of a derived graph [45].

For more on the topic, we refer to the books on graph drawing [10, 43] and the survey about orthogonal graph drawing [15].

2.2 Rectangular Dual

Let $F(R)$ be the set of rectangles of a rectangular dissection R. It is convenient to include the enclosing rectangle in $F(R)$. The *dual* of R is the graph $G_*(R)$ with vertex set $F(R)$ and edges joining pairs of rectangles that share a boundary segment; Fig. 3 shows an example. If a graph G allows a representation as dual of a rectangular dissection R, i.e., $G = G_*(R)$, then G is called a *rectangular dual* of R. It is tempting to think that the graph $G_*(R)$ is the dual graph of $G_{\mathsf{skel}}(R)$. This is almost true, but there are some issues about the multiplicity of edges incident to the outer face of $G_{\mathsf{skel}}(R)$, i.e., to the enclosing rectangle.

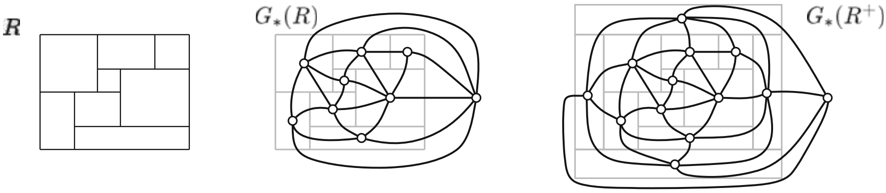


Fig. 3 A rectangular dissection R, its dual $G_*(R)$, and the dual of the framed dissection R^+

Indeed, graphs admitting a rectangular dual may have unwanted features; e.g., a vertex represented by a corner rectangle may have degree 3 and there may be a double-edge between an inner vertex w and the outer vertex v_∞. A clean characterization is obtained if we assume that the degree of v_∞ is 4. In terms of a rectangular dissection, this can be achieved by adding a frame of four rectangles, one for each side; see Fig. 3.

Theorem 2.2. *A planar triangulation with designated outer vertex v_∞ of degree 4 admits a rectangular dual exactly if it has no separating triangle, i.e., if it is 4-connected.*

There are many related characterizations, e.g., Kozmiński and Kinnen [33] or the earlier result of Ungar [59] in the dual setting. Buchsbaum et al. [2] have many pointers to the literature. An elegant approach to proving the theorem is to split the task into two: In the first step, it is shown that the graph can be enriched with some combinatorial structure. In the second step, this structure is used to construct the geometric representation. Such an approach was taken by Bhasker and Sahni [5] and later refined by He [25], Kant and He [31], and Fusy [22]. The latter two of these papers use transversal structures (cf., Sect. 5.1) as the intermediate combinatorial structure. The approach results in the linear-time construction of rectangular duals with integer coordinates bounded by n.

Problems where some region is to be partitioned into subregions subject to restrictions on the shapes of the subregions and some adjacency constraints are denoted as *floor-planning* problems. They arise in applications in VLSI chip design and cartography. In view of these applications, specific optimization tasks are of interest. We mention two directions:

- Find a floor plan of a general planar graph such that the shapes of the modules representing the vertices are simple (e.g., orthogonal with ≤ 8 corners) and the total area of the floor plan is small. This problem is studied in [37].
- A rectilinear cartogram is a diagram in which geographic regions have been replaced by orthogonal polygons. The neighbor relation on polygons and on their corresponding regions has to be the same; in addition, the areas of the polygons correspond to some numerical data associated with the regions. Eppstein et al. [16] have studied cartograms where all polygons are rectangles and with the flexibility that they can accommodate arbitrary area assignments.

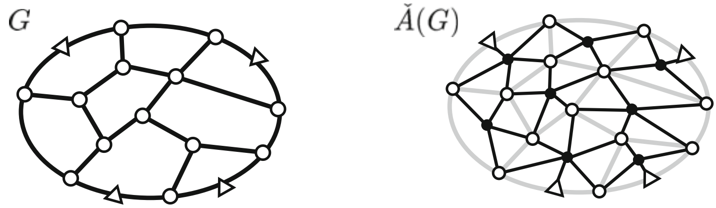


Fig. 4 A planar graph G and its trimmed angle graph $\check{A}(G)$

Alam et al. [1] show that a planar triangulation with area assignments can be represented by a cartogram using polygons with 8 corners, which is the best possible.

2.3 An Algorithm for Rectangular Drawings

Let G be the input graph. We assume that G is given with a planar embedding and with four distinguished corner vertices of degree 2; all the other degrees are 3 or 4. With $G = (V, E)$, we consider its *trimmed angle graph* $\check{A}(G)$. The vertex set of this graph consists of the primal vertex set V together with the dual vertex set save the dual of the unbounded face, i.e., $V^* \setminus \{f_\infty\}$. Edges of $\check{A}(G)$ correspond to incidences between vertices and bounded faces of G or equivalently to the internal angles of G. To emphasize the bipartition of $\check{A}(G)$, think of vertices from V as white and of vertices corresponding to faces of G as black vertices; see Fig. 4.

If G has a rectangular drawing, i.e., $G = G_{\mathsf{skel}}(R)$ for some rectangular dissection R, then we can classify the angles of G as either being a corner or being flat with respect to R. We note the following:

- Each inner face is represented as a rectangle and thus has exactly four corner angles.
- Each inner vertex of degree 3 has exactly one flat angle.

Orient the edges of $\check{A}(G)$ such that $\{v, f\}$ is oriented as $f \to v$ when v is a corner of the rectangle corresponding to f and as $v \to f$ when the angle is flat. Now consider the out-degrees of this orientation and note that

- $\mathsf{out\text{-}deg}(f) = 4$ for all black vertices f. For a white vertex v, we have $\mathsf{out\text{-}deg}(v) = 1$ if v is an inner vertex of G with $\deg(v) = 3$ and $\mathsf{out\text{-}deg}(v) = 0$ if $\deg(v) = 4$ or if v is a vertex of the outer face of G.

An orientation of $\check{A}(G)$ obeying the above rules for the out-degrees is denoted an α_{skel}-*orientation*.

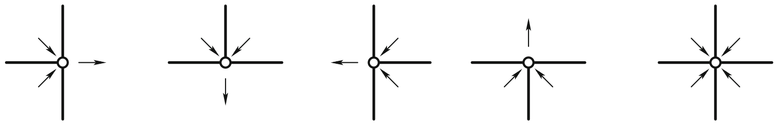


Fig. 5 Consistent alignment of rectangles sharing a vertex v; the arrows indicate the underlying α_{skel} orientation of $\check{A}(G)$

Theorem 2.3. *Let G be a plane graph with four distinguished corner vertices of degree 2 at the outer face and vertices of degrees 3 or 4 otherwise. There is a rectangular drawing of G if and only if $\check{A}(G)$ has an α_{skel}-orientation.*

Proof. From the above we know that if $G = G_{\mathsf{skel}}(R)$ for some R, then there is an α_{skel}-orientation of $\check{A}(G)$.

For the converse suppose that $\check{A}(G)$ has an α_{skel}-orientation. Identify the four corners of a rectangular frame F with the degree-2 vertices of G in clockwise order. From α_{skel}, we can read off which vertices of G are corners of a given face f; they are the out-neighbors of f in the orientation. For faces f sharing an edge with the outer face f_∞ of G which is represented by F, we know which corner is top left, top right, bottom right, and bottom left, i.e., the *alignment* of the rectangles R_f. If we know the alignment of R_f and if faces f and f' share an edge in G, then we know the alignment of $R_{f'}$. Thus, the alignment can be passed through dual paths to all faces. The claim is that the alignment of R_f is independent of the dual path from f_∞ to f that has been used. This can be established with a homotopy-type argument. The key to the argument is to check that if v is a vertex and f, f' are faces incident to v in G, then passing the alignment information from R_f to $R_{f'}$ on either of the two paths on the dual cycle around v yields the same result. This amounts to looking at the pictures of Fig. 5 with all possible choices for f and f'. □

The alignment of the rectangles is equivalent to a red–blue coloring and orientation of edges of G such that red edges are horizontal with orientation from left to right and blue edges are vertical and oriented downward. The boundary of each face consists of two directed paths in this orientation. The coloring of one of the two paths has a sequence of red edges followed by a sequence of blue edges; this is the *upper path*. The other path has blue edges followed by red edges and is called the *lower path*.

We use the red–blue coloring in the following description of how to construct the rectangular dissection R for G. Let p_0 be the lower path of the outer face f_∞. Match the blue part of p_0 to the left side of the frame F and the red part of p_0 to the bottom side of F. This requires an arbitrary specification of positions for the vertices of degree 3 contained in p_0. The third edge of each such vertex will have to be extended into the interior of F. From the coloring we know whether it is horizontal or vertical, but we do not yet know its length. Such an initial piece of an edge will be called a *stump*. Starting from $p = p_0$ we add rectangles one by one, always keeping the invariant that the boundary of the set of already placed rectangles

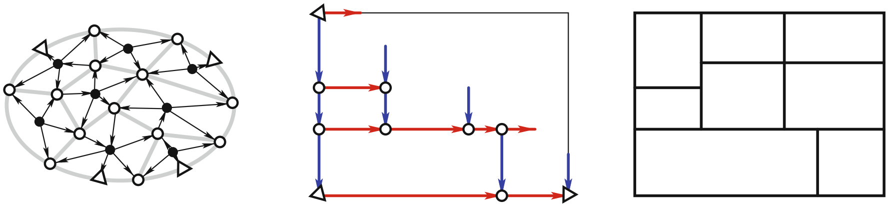


Fig. 6 An α_{skel}-orientation and a corresponding rectangular dissection

is a directed path p from the top left corner to the bottom right corner of F. We now focus on the placement of a new rectangle. Figure 6 shows an α_{skel}-orientation and an intermediate stage of the algorithm.

Along the path p we see some stumps: The first is a red stump on the top side of F, and the last is a blue stump on the right side of F. Therefore, somewhere along p there is a red stump at a vertex v_1 followed by a blue stump at v_2. Let p' be obtained by restricting p to the part between v_1 and v_2. We claim that p' consists of a sequence of blue edges followed by a sequence of red edges. This can be verified as follows: The stump at v_1 is outgoing and red, and the edge of p' incident to v_1 is also outgoing and hence must be blue. Similarly, the edge of p' incident to v_2 has to be red. A vertex of p' where a red edge is ingoing and a blue is outgoing would have a stump, which is impossible since the stumps at v_1 and v_2 are consecutive. This completes the proof of the claim. It follows that p' is the lower path of some rectangle R_f. We place R_f in the unique consistent way inside F and replace p' in p by the upper path of R_f. We also have to choose positions for the noncorner vertices contained in the upper path of R_f. Unless the top right corner of R_f is the top right corner of F and the dissection is complete, there is at least one new stump at the top right corner of R_f and possibly some more along the upper path. Now the status of the directed path p and its stumps is as before, and so the next rectangle can be placed. □

Algorithmically, the construction of $\check{A}(G)$ from a given G and the construction of the rectangle dissection R from a given α_{skel}-orientation of $\check{A}(G)$ are both easy and can be done in linear time. The computation of an α_{skel}-orientation can be modeled as a flow problem [18] in $\check{A}(G)$ and with methods from [41] be solved in $O(n^{1.5})$. In [58] it is shown that the computation of an α_{skel}-orientation can be reduced to a shortest-path problem. Using the currently fastest algorithm for planar shortest paths [42] yields an overall running time of $O(n \log^2 n / \log\log n)$. If the input graph has no vertices of degree 4, we can do even better: We construct a suitably modified dual G_q^* of G having four vertices corresponding to the outer face of G, one for each segment between degree-2 vertices on the outer face. The α_{skel}-orientations of $\check{A}(G)$ are in bijection with transversal structures (a.k.a. regular edge labelings) of G_q^* as defined by Fusy [22, 23]. In his thesis [22], Fusy showed that if G_q^* has no separating triangle, then a transversal structure exists and can be computed in $O(n)$ time. This gives an α_{skel}-orientation of $\check{A}(G)$ in linear time. The result is summarized in the following.

Theorem 2.4. *Let G be a plane graph with four distinguished corner vertices of degree 2 and vertices of degrees 3 or 4 otherwise. There is an algorithm based on* α_{skel}*-orientations that decides whether G is a skeleton graph and if so computes a rectangular drawing in* $O(n)$ *time if* $\Delta \leq 3$ *and in* $O(n \log^2 n / \log\log n)$ *time if there are vertices of degree 4.*

Rahman et al. [44] have a linear-time algorithm for rectangular drawings of graphs with maximum degree $\Delta \leq 3$ even in the case where no plane embedding of the graph is prescribed.

Miura et al. [40] have an $O(n^{2.5}/\log n)$ algorithm to recognize whether a plane graph G with $\Delta = 4$ has a rectangular drawing. Their result is stated in a more general form: They allow an outer face of more complex shape. Since they prescribe which outer vertices are convex (resp., concave) corners, the shape of the outer face can be modeled by adapting the out-degrees of α_{skel}. Therefore, this generalization is also covered by our approach. This yields a solution with $O(n \log^2 n / \log\log n)$ running time.

3 Segment Contact and Visibility Graphs

3.1 Segment Contact Graphs

Think of a rectangular dissection R as a set of segments, some horizontal and some vertical. If R contains no point where four rectangles meet, intersections between segments only occur between horizontal and vertical segments and they involve an endpoint of one of the segments; i.e., they are contacts. Otherwise, we break one of the two segments of each crossing point into two to get a system of interiorly disjoint segments. The *segment contact graph* $G_{\mathsf{seg}}(R)$ of a rectangulation R is the bipartite planar graph whose vertices are the segments of R and whose edges correspond to contacts between segments. From Fig. 7 we see that $G_{\mathsf{seg}}(R)$ is indeed planar and that the faces of $G_{\mathsf{seg}}(R)$ are in bijection with the rectangles of R and are uniformly of degree 4. Hence, $G_{\mathsf{seg}}(R)$ is a maximal bipartite planar graph, i.e., a quadrangulation.

If H is some subgraph of $G_{\mathsf{seg}}(R)$, then H can also be represented as segment contact graph of some set of interiorly disjoint horizontal and vertical segments in the plane; i.e., H is a segment contact graph. A segment contact representation for H is obtained from R by removing some segments (vertex deletion) and slightly pulling back the ends of some segments to get rid of contacts (edge deletion). The next theorem states that the converse also holds.

Theorem 3.1. *Every planar bipartite graph H admits a contact representation with interiorly disjoint horizontal and vertical segments.*

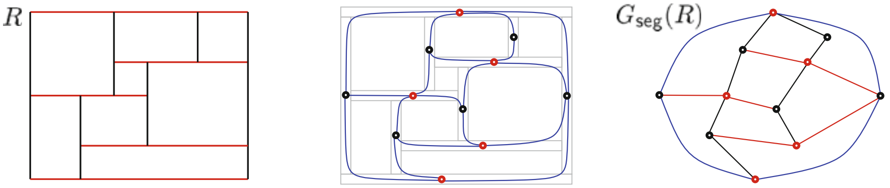


Fig. 7 A rectangular dissection R and two drawings of its segment contact graph $G_{\text{seg}}(R)$

This was shown by Hartman et al. [27] and by de Fraysseix et al. [12]. In the next subsection we sketch a proof of the theorem based on the concept of separating decompositions of quadrangulations.

A contact representation with interiorly disjoint segments such that the intersection of any $k+1$ segments is empty is called a k-contact representation. Thomassen characterized the class of graphs admitting a 2-contact representation.

Theorem 3.2. *A planar graph $G=(V,E)$ has a 2-contact representation if and only if $|E[W]| \leq 2|W|-3$ for every subset W of the vertices. As usual, $E[W]$ denotes the set of edges with both ends in W.*

Thomassen presented the result at the Graph Drawing Conference 1993 but never published his proof. Below we provide a simple proof based on rigidity theory. The condition stated in the theorem can be efficiently checked; see, e.g., Lee and Streinu [39]. Hliněný [26] showed that the recognition of general contact graphs of segments is NP-complete. Actually, he showed that even the recognition of graphs admitting a 3-contact representation is NP-complete.

A related class of graphs is the intersection graphs of segments. A longstanding conjecture dating back to Scheinerman's Ph.D. thesis was that every planar graph is a segment intersection graph. The conjecture was finally resolved by Chalopin and Gonçalves [8]. Kratochvíl and Kuběna [34] asked whether all complements of planar graphs are segment intersection graphs.

Proof of Theorem 3.2. The necessity of the condition is easily seen: Let SS be the set of segments of a 2-contact representation of G. For $W \subset V$, let X_W be the set of endpoints of segments in SS corresponding to vertices of W. We have $|X_W| = 2|W|$. There is an injection ϕ from edges in $E[W]$ to points in X_W. Points belonging to the convex hull of X_W, however, cannot be in the image of ϕ. Since the convex hull contains at least three points, we get $|E[W]| \leq |X_W| - 3 = 2|W| - 3$. $\square$

For the converse, we need some prerequisites. A *Laman graph* is a graph $G = (V,E)$ with $|E| = 2|V| - 3$ and $|E[W]| \leq 2|W| - 3$ for all $W \subset V$. Laman graphs are of interest in rigidity theory; see, e.g., [17, 24]. Laman graphs admit a *Henneberg construction*, i.e., an ordering $v_1, \dots, v_n$ of the vertices such that if G_i is the graph induced by the vertices $v_1, \dots, v_i$, then it holds that G_3 is a triangle and G_i is obtained from G_{i-1} by one of the following two operations:

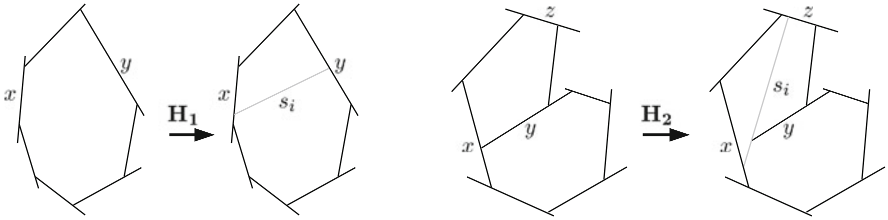


Fig. 8 The addition of segment s_i

(H_1) Choose vertices $x \neq y$ from G_{i-1} and add v_i with the two edges (v_i, x) and (v_i, y).

(H_2) Choose an edge (x, y) and a third vertex z from G_{i-1}, remove (x, y), and add v_i together with the three edges (v_i, x), (v_i, y), and (v_i, z).

In [28] it is shown that planar Laman graphs admit a planar Henneberg construction in the sense that the graph is constructed together with a plane straight-line embedding and vertices stay at their position once they have been inserted. Moreover, the Henneberg construction can start with an outer triangle that remains unchanged throughout the construction.

Now let G be a planar graph fulfilling the condition of the theorem. We may assume that G is Laman since we can easily get rid of edges in a segment contact representation by retracting ends of segments. Consider a planar Henneberg construction $G_3, \ldots, G_n = G$. Starting from three pairwise touching segments representing G_3, we add segments one by one. For the induction we need the following invariant:

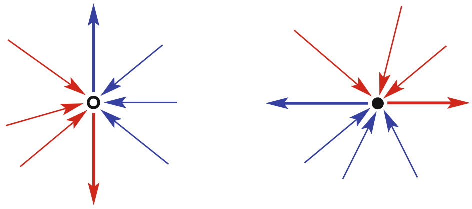
- After adding the ith segment s_i, we have a 2-contact representation of G_i and there is a correspondence between the cells of the segment representation and the faces of G_i that preserves edges; i.e., if (x, y) is an edge of the face, then one of the corners of the corresponding cell is a contact between s_x and s_y.

From the count of edges, it follows that the endpoints of all the segments except the three outer segments are used in contacts. Therefore, all faces in the segment contact representation are convex. Figure 8 shows how to add segment s_i in the cases where v_i is added by H_1 (resp., H_2). It is evident that the invariant for the induction is maintained. □

3.1.1 Separating Decompositions and Segment Contact Representations

Let Q be a quadrangulation. We call the color classes of the bipartition white and black and name the two black vertices on the outer face s and t. A *separating decomposition* of Q is an orientation and coloring of the edges of Q with colors red and blue such that

Fig. 9 Edge orientations and colors at *white* and *black* vertices



1. All edges incident to s are ingoing red and all edges incident to t are ingoing blue.
2. Every vertex $v \neq s,t$ is incident to a nonempty interval of red edges and a nonempty interval of blue edges. If v is white, then, in clockwise order, the first edge in the interval of a color is outgoing and all the other edges of the interval are incoming. If v is black, the outgoing edge is the last one in its color in clockwise order (see Fig. 9).

Separating decompositions have been studied in [11, 19, 20]. To us they are of interest because of the following lemma.

Lemma 3.3. *A segment contact representation of Q with horizontal and vertical segments induces a separating decomposition of Q.*

Proof. We assume w.l.o.g. that the segment contact representation and the plane embedding of Q are compatible in the sense that for every vertex v, the clockwise order of the edges around v corresponds to the clockwise order of the contacts around segment s_v of v. We also assume that the segments representing s and t are horizontal, s is the bottom and t is the top segment, and their endpoints have no contact with another segment.

An edge of Q is represented by a contact where an endpoint of one segment is touching the interior of another segment. Orient the edge such that the vertex contributing the endpoint is the tail of the oriented edge. This yields a 2-orientation of Q, i.e., an orientation where every vertex except s and t has out-degree 2. Since s is horizontal, all neighbors of s have to be vertical. Tracing this kind of argument through the graph, we conclude that all black vertices are represented by horizontal segments and all white vertices by vertical segments. Color the edge corresponding to the left contact of a horizontal segment blue and the edge of a right contact red. Similarly, the edge induced by the top contact of a vertical segment is blue and the edge of the bottom contact is red. This construction yields a separating decomposition of Q. For an example, see Fig. 10. □

In the following, we sketch a construction for the converse. We start with a separating decomposition of Q and construct a segment contact representation. The algorithm behind the construction may not be the fastest and the construction itself not the most flexible tool for further applications. This author has decided to include it because it nicely and unexpectedly combines some combinatorial structures. Details can be found in [19]. To begin, we need some facts.

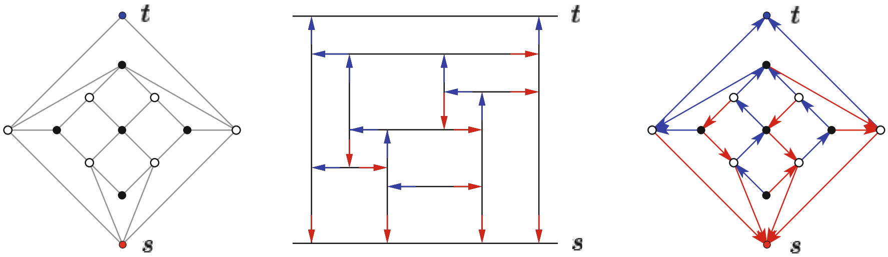


Fig. 10 A quadrangulation Q, a segment contact representation of Q, and the induced separating decomposition of Q

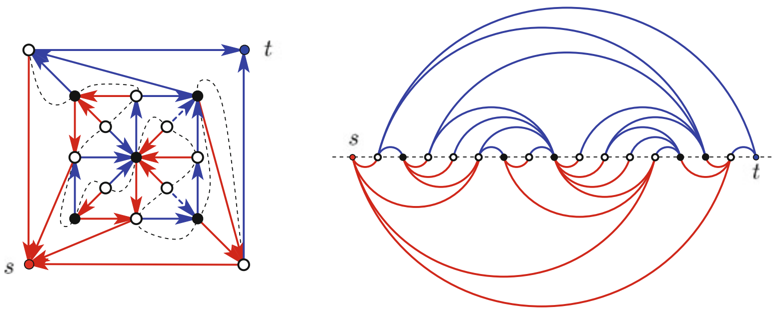


Fig. 11 A quadrangulation Q with a separating decomposition S, and the 2-book embedding induced by the equatorial line of S

- Every quadrangulation admits a separating decomposition.
- The red edges of a separating decomposition form a tree rooted at s that spans all vertices except for t. Symmetrically, the blue edges form a tree rooted at t that spans all vertices except for s.
- There exists a simple closed curve that contains all vertices of Q except s and t and avoids all edges of Q such that all red edges are in the interior of the curve and all blue edges are in the exterior. Deleting the piece of this curve that runs in the outer face, we obtain the *equatorial line* of the separating decomposition.
- By straightening the equatorial line, we obtain a 2-book embedding of Q; see Fig. 11.

An *alternating layout* of a plane tree T is a noncrossing drawing of T such that the vertices are placed on the x-axis and all edges are embedded in the half-plane above the x-axis (or all below). Moreover, for every vertex v, it holds that all its neighbors are on one side; either they are all left of v or all right of v.

It can bee shown that the 2-book embedding of Q obtained from S yields alternating layouts of the two trees of S. The roots of the two trees are the extreme vertices. In addition, black vertices have all their blue neighbors on the left and all their red neighbors on the right, while for white vertices the converse holds.

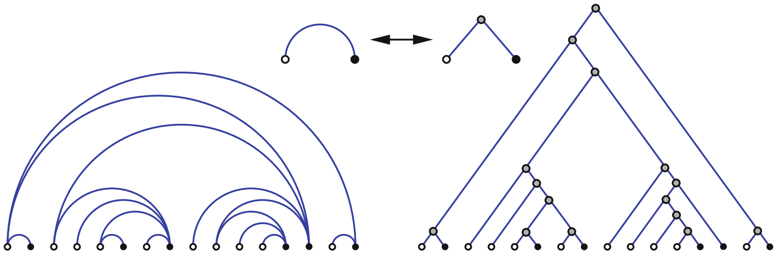


Fig. 12 A bijection between alternating trees and the binary trees

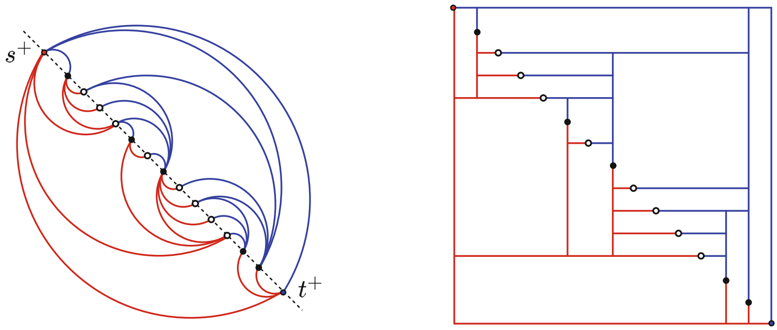


Fig. 13 The 2-book embedding from Fig. 11 after merging s and t with their respective neighbors and the associated rectangulation

Figure 12 indicates a bijection between alternating trees and binary trees such that left/right vertices of the alternating tree correspond to left/right leaves of the binary tree.

Modify the 2-book embedding by merging s with its successor into a new vertex s^+ and symmetrically t with its predecessor into t^+. Then apply the bijection of Fig. 12 to each of the two trees on its side and tilt the picture by 45°. The result is a segment contact representation of Q; see Fig. 13.

3.2 Visibility Graphs

A family of disjoint horizontal segments in the plane defines a *visibility graph*. The vertices of the graph are the segments and edges are based on vertical visibility: A segment s' is *visible* from segment s if there is a vertical ray r leaving s such that s' is the first segment in the family reached by r. The visibility graph is undirected since if s' is visible from s via an up-ray, then there is a down-ray proving visibility of s from s'. However, if we only care about the up-rays, we get a

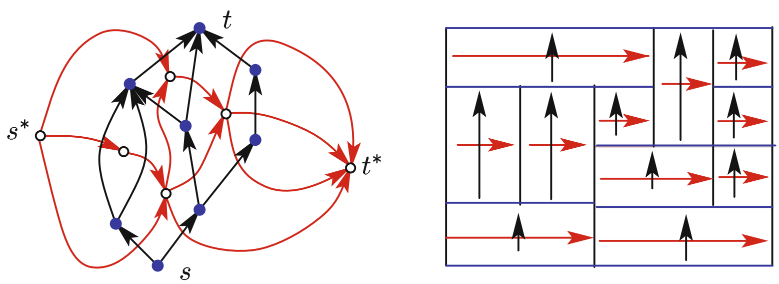


Fig. 14 A dual pair B, B^* of bipolar orientations and the rectangular dissection returned by the algorithm when using B and B^*

natural orientation of the visibility graph; this orientation is a bipolar orientation. Indeed, bipolar orientations and visibility representations of planar graphs are intimately related. The study of this connection and its use for the construction of visibility representations has been pioneered by Rosenstiehl-Tarjan [47] and Tamassia-Tollis [55].

3.2.1 Bipolar Orientations

Let G be a graph with distinguished adjacent vertices s and t. A *bipolar orientation* of G is an acyclic orientation of G such that s is the unique source and t is the unique sink; the oriented edge (s,t) is the *root* of the orientation. A graph G admits a bipolar orientation exactly if G is 2-connected. Actually, the root edge of a 2-connected graph can be chosen arbitrarily. In the literature these facts are frequently treated in the disguise of st-numberings. Sometimes it is convenient to omit the root edge. In a slight abuse of notation, we also speak of a bipolar orientation in this case. This is done, for instance, in the next paragraph and in Fig. 14.

Bipolar orientations of a plane 2-connected (multi)graph G with designated s and t on the outer face are characterized by two facts:

Fact F. Every face f of G has exactly two angles where the orientation of the edges coincide. [The vertices incident to these angles are $v_{\mathsf{source}}(f)$ and $v_{\mathsf{sink}}(f)$.]
Fact V. Every vertex $v \neq s,t$ of G has exactly two angles where the orientation of the edges differ. [The faces incident to these angles are $f_{\mathsf{sink}}(v)$ and $f_{\mathsf{right}}(v)$.]

An orientation of a plane graph G induces an orientation on the dual graph G^*: Define the orientation of a dual edge e^* as left to right relative to the orientation of e. That is, when looking in the direction of an oriented edge e, the dual edge e^* is oriented from the left face to the right face of e.

The *bipolar dual* of a plane bipolar orientation is this dual graph endowed with this dual orientation rooted at (s^*,t^*), where s^* is be the face on the right and t^* is the face on the left of (s,t); i.e, the orientation of (s^*,t^*) is not dual to (s,t). Essentially, facts V and F are dual and we have set up the orientation of the root edge (s^*,t^*) such that the bipolar dual of a plane bipolar orientation is again a bipolar orientation.

A more comprehensive treatment, including proofs of the material in this subsection, can be found in [13], for example.

3.2.2 From a Bipolar Orientation to a Rectangular Dissection

The input of the following algorithm is a 2-connected plane graph G with an oriented root edge (s,t) on the outer face. The output is a rectangular dissection R such that the visibility graph of the horizontal segments of R is G.

Algorithm Rectangular Dissection

- Compute a bipolar orientation B of G with root edge (s,t).
- Compute the bipolar dual B^* with root edge (s^*,t^*) of B.
- For each primal vertex v, let $y(v)$ be the length of the longest directed $s \to v$ path in B.

  For each dual vertex f, let $x(f)$ be the length of the longest directed $s^* \to f$ path in B^*.
- With a primal vertex $v \neq s,t$, associate the horizontal segment with left end at the point $(x(f_{\mathsf{left}}(v)),y(v))$ and right end at $(x(f_{\mathsf{right}}(v)),y(v))$. With a dual vertex $f \neq s^*,t^*$, associate the vertical segment with lower end $(x(f),y(v_{\mathsf{source}}(f)))$ and upper end $(x(f),y(v_{\mathsf{sink}}(f)))$.

  The special vertices s, t, s^*, and t^* need special treatment; as endpoints of these four segments we can choose the four points $(0,0)$, $(0,y(t))$, $(x(t^*),y(t))$, and $(x(t^*),0)$. If the visibility between s and t is required, the right endpoint of these two segments can be shifted to the right by one unit.

From the algorithm, we obtain

Theorem 3.4. *For a 2-connected planar graph G with n vertices and n^* faces, there is a visibility representation with horizontal segments whose endpoints are on integer points (x,y) with $0 \leq x \leq n^*$ and $0 \leq y \leq n-1$. Such a representation can be computed in linear time.*

A sketch of the proof of correctness for the algorithm can be found in [10]. A sweep-like proof may, however, be simpler. This could be accomplished by an inductive proof of the statement that the set of segments with y-coordinate $\leq k$ is a visibility representation of the graph induced by vertices v with $y(v) \leq k$.

The bound on the size of the representation accounted for a visibility between s and t.

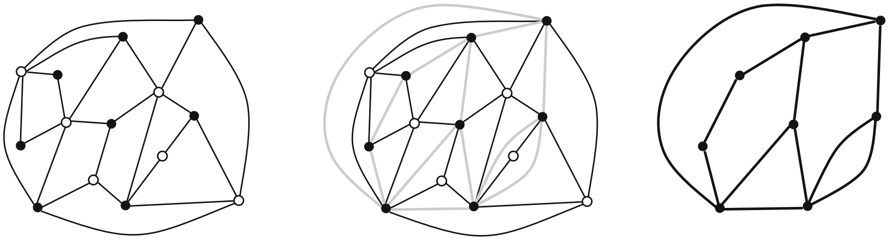


Fig. 15 A quadrangulation Q and its black graph G_b

Quite some research has been put into more compact visibility representations. The basic idea is that a carefully chosen bipolar orientation may have a longest $s \to t$ path of length much less than n. Almost optimal bounds are known:

- Zhang and He [60] show that there are planar graphs on n vertices requiring size at least $(\lfloor \frac{2n}{3} \rfloor) \times (\lfloor \frac{4n}{3} \rfloor - 3)$ for a visibility representation.
- He and Zhang [29] show that every planar graph with n vertices has a visibility representation with height at most $\frac{2n}{3} + 2\lceil \sqrt{n/2} \rceil$. They use properties of Schnyder woods to construct an appropriate bipolar orientation.
- Fan et al. [21] show that every planar graph with n vertices has a visibility representation with width at most $\lfloor \frac{4n}{3} \rfloor - 2$.
- Sadasivam and Zhang [52] show that it is NP-hard to find the bipolar orientation of G which minimizes the length of the longest $s \to t$ path.

3.3 Bipolar Orientations and Separating Decompositions

Let again Q be a plane quadrangulation with color classes consisting of black and white vertices. The *black graph* G_b of Q is the graph on the set V_b of black vertices of G, where $u, v \in V_b$ are connected by an edge for every face f of Q incident to u and v; i.e., there is a bijection between faces of Q and edges of G_b. The graph G_b inherits a plane embedding from the plane embedding of Q. Note that in general G_b may have multiple edges.

Let G be a plane graph; the *angle graph* of G is the graph Q with vertex set consisting of vertices and faces of G, and edges corresponding to incidences between a vertex and a face, and Q inherits a plane embedding from G. If there are no multiple incidences, i.e., if G is 2-connected, then Q is a quadrangulation. The angle graph construction $G \to Q_G$ is the inverse of the black graph construction $Q \to G_b$. More precisely, $Q \leftrightarrow G_b$ is a bijection between plane quadrangulations with a black-white coloring and plane 2-connected multigraphs. An example is given in Fig. 15.

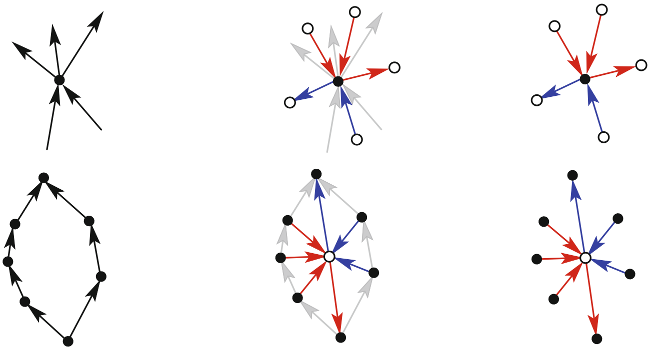


Fig. 16 From a bipolar orientation to a separating decomposition and back

Changing the role of the color classes we can associate the *white graph* G_w with Q. By symmetry, $Q \leftrightarrow G_w$ is again a bijection between plane quadrangulations Q with a black-white coloring and plane 2-connected multigraphs. The induced bijection $G_b \leftrightarrow G_w$ is nothing more than the traditional plane duality.

The study of the oriented counterpart of these classical bijections goes at least back to [47] and was continued in [11,13,19,46]. Let B be a plane bipolar orientation on black vertices with the root edge (s,t) on the outer face. Let Q be the angle graph of the underlying undirected graph of B. Facts V and F of bipolar orientations yield two special angles for every $v \neq s,t$ of B and for every face of B. At vertices, the special angles are the *left* and the *right* angle. At faces, the special angles are the *source* and the *sink* angle. Using the correspondence between angles of B and edges of Q, we define a separating decomposition on Q: The edge incident to v in Q that corresponds to the left special angle is outgoing blue, and the edge that corresponds to the right special angle is outgoing red. The edge incident to f in Q that corresponds to the source is red outgoing and the edge corresponding to the sink is blue outgoing. The rules are illustrated in Fig. 16. It is easily verified that they yield a separating decomposition of Q as defined in Sect. 3.1.1.

Starting from a separating decomposition S on Q, we obtain the unique bipolar orientation B on G_b inducing S by using the converse rules: At a vertex v, the two outgoing edges of S split the edges of G_b into two blocks. The block where Q may have blue edges is the block of incoming edges in the bipolar orientation, and the edges of the other block are the outgoing edges in the bipolar orientation. The oriented bijection fits together well with oriented duality; there is only a slight asymmetry concerning the outer vertices and edges.

The correspondence between bipolar orientations and separating decomposition allows us to use the construction from Sect. 3.1.1 for visibility representations and the algorithm from Sect. 3.2.2 for segment contact representations.

4 Square Dissections

In this section, we are mainly interested in *square dissections*, which are rectangular dissections where all the small rectangles are squares. Again, we are going to look at different graphs associated with such a dissection.

4.1 Squarings and Electricity

The theory of square layouts goes back to a seminal paper, ”The dissection of rectangles into squares,” by Brooks, Smith, Stone and Tutte from 1940 [7]. The story of the collaboration of the four students has been told several times. In [57] Tutte tells his memories. The aim of the four students was to find squarings of a square such that the sizes of all small squares are different. Such a squaring is called *perfect*; see Fig. 17.

They based the search for perfect squarings on the following idea: Start with a rectangular dissection and try to produce a combinatorially equivalent squaring.

The squaring process could be simplified with some observations: The initial rectangular dissection R can be described by the visibility graph G of the horizontal segments. The rectangles of the dissection are in bijection with the edges of G (this may require multiple edges in G). On G there is the natural upward- pointing bipolar orientation B [the root edge (s,t) can be omitted].

With an arc a of B, associate the width $w(a)$ of the rectangle corresponding to a in R. The function w on the arcs of B respects the flow conservation law in every vertex except s and t; it is an st-flow.

Symmetrically, from the dual graph G^* with the dual bipolar orientation B^* and the height $h(a^*)$ of the rectangle associated with a^*, an s^*t^*-flow is obtained.

If the dissection R were a squaring, then $w(a) = h(a^*)$ for every dual pair $a \leftrightarrow a^*$ of arcs.

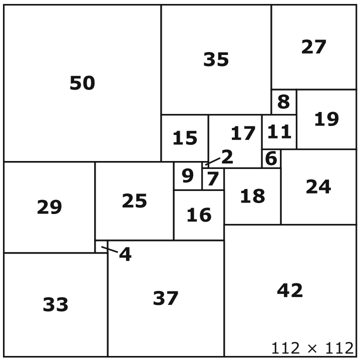


Fig. 17 The essentially unique perfect squaring with the smallest number of squares. It was discovered in 1978 by A.J.W. Duijvestijn

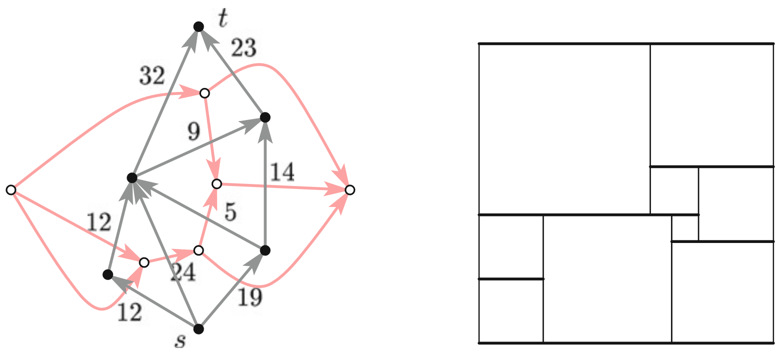


Fig. 18 A flow with primal and dual flow conservation and the corresponding squaring

It can be verified that if we find a function f on G such that with $f(e^*) = f(e)$, we have an st-flow on G and an s^*t^*-flow on the dual G^*, then orienting the edges such that $f(a) \geq 0$ yields a bipolar orientation on G. Based on this bipolar orientation, a squaring of G such that f describes the size of the squares can be constructed with a weighted variant of the algorithm from Sect. 3.2.2. An example is shown in Fig. 18.

Physicists may recognize a flow with primal–dual flow conservation as an electrical flow. Primal and dual flow conservation corresponds to the two Kirchhoff laws. To find an appropriate f, we can thus build G as an electrical network such that each edge has unit resistance and then attach the poles of a battery to s and t and measure the electrical current in each edge. The solution can, of course, also be obtained analytically. In fact, Kirchhoff's theorem [32] relates the solution to the enumeration of certain spanning trees.

Theorem 4.1. *Let G be a plane graph with special vertices s and t on the outer face. For an edge $e = \{u,v\}$ of G, let $w_e = |n_{uv} - n_{vu}|$, where n_{xy} is the number of spanning trees T of G such that T contains a path $s,..,x,y,..,t$. If $w_e > 0$ for all e, then there is a square layout R with G as visibility graph such that the square associated with e has side length w_e.*

Sketch of a proof. We use the duality $T \leftrightarrow T^*$ between spanning trees of G_b and its dual $G_b^* = G_w$. If a tree T contributes to n_{uv}, then T^* contains the edge $\{s^*,t^*\}$. Let f_l and f_r be the faces left and right of (u,v) in G_b, and define $S = T - \{u,v\} + \{s,t\}$ and $S^* = T^* - \{s^*,t^*\} + \{f_l, f_r\}$. The pair S, S^* is again a dual pair of spanning trees. The mapping $(T,T^*) \leftrightarrow (S,S^*)$ yields a bijection between pairs contributing to n_{uv} and pairs contributing to the dual count $n^*_{f_l f_r}$. From the existence of this bijection, it can be concluded that $w_{\{u,v\}} = w_{\{f_l,f_r\}}$. □

Our proof is taking advantage of the planarity of G. However, for general graphs, the same definition of w_e yields the electrical current in e (up to normalization); see, e.g., [4].

If $w_e = 0$ for some edge e, we still get a square layout R. In R the horizontal segments corresponding to the two end vertices of e are merged into a single

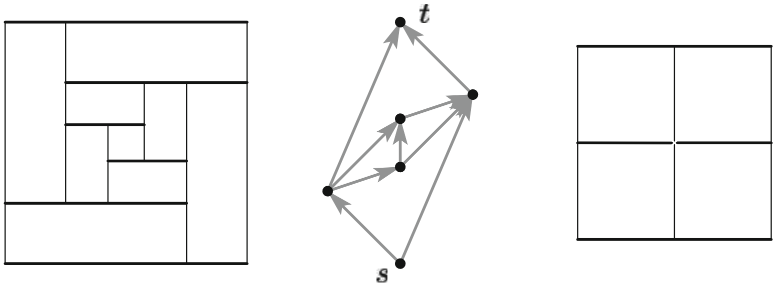


Fig. 19 A rectangular dissection R, the bipolar orientation of the visibility graph G obtained from R and the squaring of G with five invisible zero squares in the center

segment. If all edges with zero flow are isolated, this may be tolerated as consistent with the notion of visibility representation; however, as shown in Fig. 19, more complex parts of a graph can disappear because all their edges have zero flow. A sufficient condition that ensures that edges with zero flow are at least isolated is that the graph is 3-connected (see Proposition 4.5).

4.2 Squarings and Separating Decompositions

We now come to a different approach to square layouts. Associate a variable with every inner face of a quadrangulation Q. In these variables we set up a system of linear equations such that a nonnegative solution of the system yields a squaring. To define the system of equations, enhance the quadrangulation Q with a separating decomposition S. We aim for a squaring that induces S with its segment contacts.

For a vertex v of Q, let $\mathcal{F}_r(v)$ be the set of faces incident to v in the angle between the two outgoing edges of v where incoming edges are red, and let $\mathcal{F}_b(v)$ be the other incident faces of v, i.e., the faces in the angle with blue incoming edges.

Suppose that such a square representation inducing S exists, and let x_a be the sidelength of the square representing a face a of Q. Every inner vertex v implies an equation that has to be fulfilled by the side lengths:

$$\sum_{a\in\mathcal{F}_r(v)} x_a = \sum_{a\in\mathcal{F}_b(v)} x_a. \tag{1}$$

A quadrangulation with n vertices has $n-2$ faces; hence, we have a system of $n-4$ linear equations (inner vertices) in $n-3$ variables (inner faces). To forbid the trivial solution of the homogenous system, we let $\mathcal{F}_r(t)$ be the set of bounded faces incident to t and add the equation $\sum_{a\in\mathcal{F}_r(t)} x_a = 1$. Rewriting the equations (1) as $\sum_{a\in\mathcal{F}_r(v)} x_a - \sum_{a\in\mathcal{F}_b(v)} x_a = 0$ and collecting all of them in a matrix A_S, we find that the vector of side lengths is a solution to the system

$$A_S \cdot x = \mathbf{e_1}. \tag{2}$$

Given any separating decomposition S of Q, we can consider the corresponding system (2). In the following, we show

- The matrix A_S is nondegenerate; hence, there is a unique solution x to the system.
- If the solution vector x is nonnegative, then there is a squaring with side lengths given by the components of x. If the solution vector x is positive, then the separating decomposition induced by the segment contacts of the squaring is S.

Theorem 4.2. *The matrix A_S is nondegenerate; i.e.,* $\det(A_S) \neq 0$.

Proof. The idea for the proof is to show that $|\det(A_S)|$ is the number of perfect matchings of an auxiliary graph H_S. Let $\hat{A}_S$ be the matrix obtained from A_S by replacing each -1 by 1 and note that $\hat{A}_S$ only depends on Q and not on S. It has a 1 for every incidence between an inner face (variable) and an inner vertex or t (equation). Regard $\hat{A} = \hat{A}_S$ as the adjacency matrix of a bipartite graph H. From what we said before, the graph H is obtained from the angle graph $A(Q)$ of Q by removing the vertices s, s^*, t^* and the vertex corresponding to the outer face of Q together with the incident edges. This implies

- H is planar.
- All inner faces of H are quadrangles.

Consider the Leibniz-expansions of $\det(A_S)$. The nonvanishing summands $\prod_i a_{i\,\sigma(i)}$ are in bijection with the perfect matchings M_σ of H. The contribution of a perfect matching M to $\det(A_S)$ is either $+1$ or -1; it will be denoted $\mathsf{sign}_S(M) = \mathsf{sign}(\pi_M)\prod_{ij\in M}[A_S]_{ij}$. □

The proof of the theorem relies on the following two claims:

Claim A. The graph H has a perfect matching.
Claim B. If M and M' are perfect matchings of H, then $\mathsf{sign}_S(M) = \mathsf{sign}_S(M')$.

We first prove Claim A by verifying the Hall condition for H. Let (X, R) be the vertex bipartition of H where X corresponds to the set of inner faces of Q and $R = V(Q) \setminus \{s, s^*, t^*\}$.

The Hall condition for the full angle graph $A(Q)$ is easily verified: Consider a set F of inner faces of Q and let Z be the set of connected components of $\mathbb{R}^2 \setminus \bigcup_{f\in F} \bar{f}$, where $\bar{f}$ is the closure of face f. The set $F \cup Z$ is the set of faces of a planar bipartite graph whose vertices and edges are those incident to elements of F in Q. This graph has at least $\frac{1}{2}(4|F| + 4|Z|)$ edges and hence, by Euler's formula at least $|F| + |Z| + 2 \geq |F| + 3$ vertices. This implies the Hall condition for H, as H equals $A(Q)$ after the removal of three vertices.

For the proof of Claim B, we need a slight extension of Claim A, namely, that every edge of H takes part in some perfect matching. This can be verified by showing that $V(Q) \setminus \{s, s^*, t^*\}$ is the unique nonempty subset of vertices of Q such that the Hall condition for the set is tight in H. Now we use some facts about α-orientations of planar graphs from [18]:

- Perfect matchings of H are in bijection with orientations of H with $\mathsf{out\text{-}deg}(x) = 1$ for all $x \in X$ and $\mathsf{out\text{-}deg}(r) = \deg(r) - 1$ for all $r \in R$. We refer to these orientations as α_M-orientations.
- It is possible to move from any α_M-orientation to any other α_M-orientation by *flips* of the following type: Select an inner face whose boundary is an oriented cycle and revert the orientation of all edges of this cycle. This is true (see [18]) because α_M-orientations have no rigid edges; this follows because every edge takes part in some perfect matching.

Consider two perfect matchings M and M' of H such that the corresponding α_M-orientations differ by a single flip. Since all the faces of H are quadrangles, the permutations π_M and $\pi_{M'}$ differ by a single transposition, whence $\mathsf{sign}(\pi_M) = -\mathsf{sign}(\pi_{M'})$. To obtain $\mathsf{sign}_S(M) = \mathsf{sign}_S(M')$, we need $\prod_{ij\in M}[A_S]_{ij} = -\prod_{ij\in M'}[A_S]_{ij}$. Since all entries of A_S are $+1$ or -1, it is enough to show that if we multiply the four entries of A_S associated with the four edges of a face of H, the result will always be -1. A face of H is a cycle of the form (x_1, r_1, x_2, r_2) with $x_i \in X$ and $r_j \in R$. In Q we have the edge $r_1 r_2$, which is oriented in the separating decomposition S, say as $r_1 \to r_2$. From the definition of the equations based on S, we find that $[A_S]_{r_1 x_1} = -[A_S]_{r_1 x_2}$ and $[A_S]_{r_2 x_1} = [A_S]_{r_2 x_2}$. From this we obtain $\prod_{ij=1,2}[A_S]_{r_i x_j} = -1$. Since we can move between any two matchings with flips and since flips leave the sign unaffected, we have proved Claim B. □

The theorem tells us that the linear system (2) has a unique solution x_S. The solution, however, need not be nonnegative. What we do next is to show that based on a solution x_S containing negative entries, we can modify S to obtain a new separating decomposition S' such that the solution $x_{S'}$ of the system corresponding to S' is nonnegative.

Consider a rectangular dissection R representing S and color gray all rectangles whose value in the solution vector x_S is negative. Let Γ be the boundary of the gray area in R. Here is a simple but useful lemma.

Lemma 4.3. *The boundary Γ contains no complete segment.*

Proof. Suppose Γ contains the complete segment corresponding to a vertex v of Q. Then we have $x_a < 0$ for all $a \in \mathcal{F}_r(v)$ and $x_a \geq 0$ for all $a \in \mathcal{F}_b(v)$ or the converse. In either case, we get a contradiction because the entries of x_S satisfy Eq. (1). □

Let s_0 be any segment that contributes to Γ. From the lemma we know that at some interior point of segment s_0 the boundary snaps off and continues on another segment s_1. Again, the boundary has to leave s_1 at some interior point to continue on s_2. Because this procedure always follows the boundary of the gray region, it has to turn back to segment s_0. Figure 20 shows an example.

Recall that a 2-orientation of a quadrangulation with white and black vertices is an orientation of the edges such that the two black vertices s and t on the outer face have out-degree 0 and all the other vertices have out-degree 2. In [11] it was shown that separating decompositions and 2-orientations of Q are in bijection. Reverting a directed cycle in a 2-orientation yields another 2-orientation.

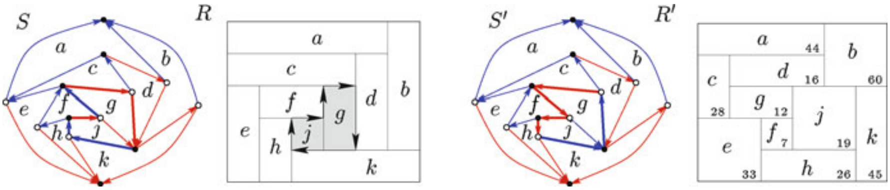


Fig. 20 The separating decomposition S corresponds to the rectangular layout R. In the solution x_S of $A_S \cdot x = \mathbf{e_1}$, the *gray rectangles* are those with a negative value. The boundary Γ consists of a single cycle whose reversal yields S'. The dissection R' corresponding to S' can be squared; small numbers are multiples of the entries of the solution $x_{S'}$

Now Γ corresponds to some directed cycles in S reverting the cycles yields another 2-orientation and via the bijection another separating decomposition S'.

Lemma 4.4. *Let S' be obtained from the separating decomposition S by reverting the cycles of the boundary Γ between faces with positive and negative x_S values. There are no negative entries in the solution vector $x_{S'}$ of the system $A_{S'} \cdot x = \mathbf{e_1}$.*

Proof. Let O be a 2-orientation of a quadrangulation Q and let C be a directed cycle in O. Let O' denote the 2-orientation obtained from O by reverting C. The relation between the separating decompositions S and S' corresponding to O and O' was investigated in [20]. There it is shown that all edges outside the cycle keep their color while all edges in the interior of the cycle change their color. □

Now consider the matrices A_S and $A_{S'}$. The rows correspond to the outer vertex t and the inner vertices of Q, and the columns correspond to bounded faces of Q. An entry $a_{v,f}$ is nonzero, more precisely $a_{v,f} = \pm 1$, if v is a boundary vertex of f. Only the sign of an entry depends on the separating decomposition; it is positive if f belongs to $\mathcal{F}_r(v)$ (the set of faces incident to v in the angle between the two outgoing edges of v where incoming edges are red) and it is negative if f belongs to $\mathcal{F}_b(v)$.

From the above it follows that if f is inside C and v is incident to f, then f belongs to $\mathcal{F}_b(v)$ with respect to S if and only if f belongs to $\mathcal{F}_r(v)$ with respect to S'. In other words, if $A_S = (a_{v,f})$ and $A_{S'} = (a'_{v,f})$, then

$$a'_{v,f} = \begin{cases} -a_{v,f} & f \text{ inside of } C \\ a_{v,f} & \text{otherwise.} \end{cases}$$

The solution $x_{S'}$ of $A_{S'} \cdot x = \mathbf{e_1}$ can, therefore, be obtained from the solution x_S of $A_S \cdot x = \mathbf{e_1}$ by changing the sign of all entries of x_S that correspond to faces f inside C.

Since S' was obtained by reversing the boundary enclosing all negative faces of x_S, it follows that $x_{S'}$ is nonnegative. □

Given a nonnegative solution, it remains to actually construct the squaring with the given sizes. We omit the details but note that the squaring can be constructed using a weighted variant of the algorithm in Sect. 3.2.2.

The electricity approach to squarings implies that to a rectangular dissection there corresponds a unique squaring. This must be the same as the squaring that we obtain by following the approach of this subsection.

There is again the issue of squares of size 0. If we color faces of S corresponding to zero squares gray and consider the boundary Γ of the gray region, then Lemma 4.3 holds again. Hence, again, a cycle of Γ is a directed cycle in S. Looking at the square representation, we see that a region of zero squares is incident to at most four nonzero squares. A cycle in Γ must therefore be of length at most 4. As a consequence, we get

Proposition 4.5. *If Q contains no separating 4-cycle, then the square dissection contains a segment contact representation of Q [we have to allow that two horizontal (resp., vertical) segments share an endpoint].*

4.3 Trapezoidal Dissections and Markov Chains

The connection between square dissections and electrical networks was used by Dehn [9] in 1903 to show that if an $A \times B$ rectangle admits a squaring using finitely many squares, then A/B is rational. Tutte used network methods in his investigations of dissections using equilateral triangles [56]. In [50] dissections into triangles are constructed using a generalized "unsymmetrical" electricity. Since there is a well-known connection between electrical networks and random walks (see e.g. [14]), it is consequent to base the construction of dissections on random walks. This approach has been taken by Kenyon [30].

For the description of Kenyon's ideas, it is convenient to start with a tiling of a rectangle into trapezoids with horizontal upper and lower sides, known as a *trapezoidal dissection*. With a trapezoidal dissection $\mathcal{T}$, associate the *t-visibility graph* $G_{\mathcal{T}}$: The vertices of $G_{\mathcal{T}}$ are the horizontal segments of the dissection, and the edges of $G_{\mathcal{T}}$ correspond to the trapezoids of $\mathcal{T}$. The lower and upper segments of the enclosing rectangle are denoted s and t. An example is shown in Fig. 21.

For a trapezoid T whose horizontal sides are on segments i and j, we let $\mathsf{height}(T)$ be the distance between segments i and j and $\mathsf{width}_i(T)$ be the length of the side of T contained in segment i; note that $\mathsf{width}_i(T) = 0$ is possible. Define an unsymmetrical weighting on edges: $m(i,j) = \frac{\mathsf{width}_i(T)}{\mathsf{height}(T)}$. These weights are used to define a random walk (Markov chain) on $G_{\mathcal{T}}$ by taking the probability $p(i,j)$ of a transition from i to j proportional to $m(i,j)$; i.e., $p(i,j) = \frac{1}{\sum_j m(i,j)} m(i,j)$.

Consider a stationary distribution π of p, that is, a distribution such that for all i:

$$\pi(i) = \textstyle\sum_j \pi(j) p(j,i).$$

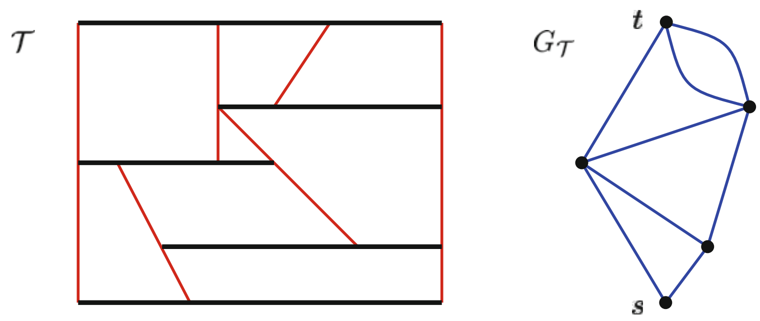


Fig. 21 A trapezoidal dissection $\mathcal{T}$ and its t-visibility graph $G_{\mathcal{T}}$

In most cases the Markov chain p induced by $\mathcal{T}$ will be aperiodic; in that case p is ergodic and π is unique. For an edge (i,j), define $\pi(i,j) = \pi(i)p(i,j)$ and note that $\pi(i,j)$ is the probability of the presence of the edge in the random walk. More formally, it is the stationary distribution for the induced random walk on the line graph of $G_{\mathcal{T}}$.

Define a function ω on the faces of $G_{\mathcal{T}}$ by

(W_0) $\omega(f_\infty) = 0$, where f_∞ is the outer face.
(W_1) $\omega(f) - \omega(f') = \pi(i,j) - \pi(j,i)$ when (i,j) is an edge with f on the left and f' on the right side.

Lemma 4.6. *The function w is well defined.*

Proof. First note that if edges (i,j) and (i',j') have f on the left and f' on the right, then the removal of the two edges cuts $G_{\mathcal{T}}$ into two components H and H' each containing at least one vertex. The random walk has to commute in both directions between these components equally often; i.e., $\pi(i,j) + \pi(j',i') = \pi(j,i) + \pi(i',j')$. Therefore, $\omega(f) - \omega(f')$ is independent of the edge chosen for the definition. □

The value of $\omega(f)$ can be determined by taking a dual path from f_∞ to f. To show that the result is independent of the path, it is enough to show that summing up the differences $\pi(i,j) - \pi(j,i)$ on a dual path around vertex i results in zero. This follows from

$$\sum_j \pi(i,j) = \sum_j \pi(j,i).$$

To prove this, note that $\pi(i) = \pi(i)\sum_j p(i,j) = \sum_j \pi(i)p(i,j) = \sum_j \pi(i,j)$, and because π is the stationary distribution, also $\pi(i) = \sum_j \pi(j)p(j,i) = \sum_j \pi(j,i)$. □

The function $\omega(x)$ is the expected relative counterclockwise winding number of the random walk around face x. Clearly, the winding number has to comply with the two properties (W_0) and (W_1), but as shown in the lemma, this determines the function.

For a face f of $G_{\mathcal{T}}$, let s_f be the corresponding line segment in $\mathcal{T}$ and define $w(f) = 1/\mathsf{slope}(s_f)$ and $w(f) = 0$ if s_f is vertical. Recall that $m(i,j) = \mathsf{width}_i(T)/\mathsf{height}(T)$.

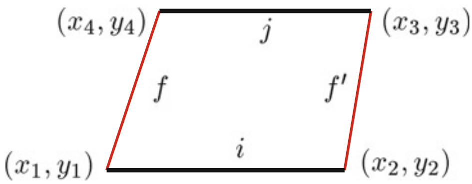


Fig. 22 An example trapezoid

Proposition 4.7. *Up to a scalar factor, the functions $w(f)$ and $m(i,j)$ defined on $\mathcal{T}$ equal the winding number $\omega(f)$ and the edge-stationary distribution $\pi(i,j)$ of the Markov chain.*

Proof. Consider a trapezoid from $\mathcal{T}$ as shown in Fig. 22. Taking the coordinates from there, we have

$$m(i,j)=\frac{x_2-x_1}{y_4-y_1}, \quad m(j,i)=\frac{x_3-x_4}{y_4-y_1}, \quad w(f)=\frac{x_4-x_1}{y_4-y_1}, \quad w(f')=\frac{x_3-x_2}{y_3-y_2}=\frac{x_3-x_2}{y_4-y_1}.$$

It follows that $w(f)-w(f')=m(i,j)-m(j,i)$. Summing up this equation along the dual cycle around a vertex, we obtain $\sum_j m(i,j)=\sum_j m(j,i)$. This implies that $m(i,j)$ is a scalar multiple of an edge-stationary distribution. In turn from (W_0) and (W_1), it follows that the values $w(f)$ are multiples of the winding numbers. □

Proposition 4.8. *For a vertex i of $G_{\mathcal{T}}$, let $y(i)$ be the y-coordinate of the corresponding segment in $\mathcal{T}$ and let $p(i,j)$ be as above. For all $i \notin \{s,t\}$, the function y is harmonic with respect to p; i.e.,*

$$y(i)=\sum_j y(j)p(i,j).$$

Proof. Let $I^+=\{j : y(j)>y(i)\}$ and $I^-=\{j : y(j)<y(i)\}$ and note that if T_{ij} is a trapezoid connecting segments i and j, then $\mathsf{height}(T_{ij})=y(i)-y(j)$ if $j\in I^-$ and $\mathsf{height}(T_{ij})=-(y(i)-y(j))$ if $j\in I^+$. We now have

$$\begin{aligned} y(i)-\sum_j y(j)p(i,j) &= \sum_j (y(i)-y(j))p(i,j) \\ &= \sum_{j\in I^-}(y(i)-y(j))p(i,j)+\sum_{j\in I^+}(y(i)-y(j))p(i,j) \\ &= \frac{1}{\sum_j m(i,j)}\Big(\sum_{j\in I^-}\mathsf{width}_i(T_{ij})-\sum_{j\in I^+}\mathsf{width}_i(T_{ij})\Big)=0. \end{aligned}$$

The last equation follows because the trapezoids with segment i on the low side and those with the segment on the high side can both be used to partition the segment. □

From the tiling $\mathcal{T}$, we have obtained a Markov chain p on $G_{\mathcal{T}}$ such that $m(i,j) \sim \pi(i,j)$ and $w(f) \sim \omega(f)$ and the y-coordinates of the vertex segments are a function that is p-harmonic for all vertices $i \notin \{s,t\}$ of $G_{\mathcal{T}}$.

To revert the process, let $G = (V,E)$ be a planar graph with s and t on the outer face and let p be a Markov chain on G. From these data, we obtain the edge-stationary distribution $\pi : E \to \mathbb{R}$, the winding numbers $\omega(f)$ on the faces of G, and the unique p-harmonic function $y : V \to \mathbb{R}$ with $y(s) = 0$ and $y(t) = 1$.

Note that the probability $h(i)$ that a random walk started at i reaches t before reaching s has the properties required for $y(i)$. Uniqueness, i.e., the fact that $h(i) = y(i)$, follows from the maximum principle for discrete harmonic functions; see, e.g., [3].

From the data, we build a trapezoidal dissection $\mathcal{T}$ such that the trapezoid T_{ij} corresponding to an edge (i,j) of G with $y(i) < y(j)$ is a horizontal translate of the trapezoid with corners

$$
\begin{aligned}
(\omega(x)(y(j)-y(i)),y(j)), \qquad & ((\pi(i,j)+\omega(x'))(y(j)-y(i)),y(j)),\\
(0,y(i)), \qquad & (\pi(i,j)(y(j)-y(i)),y(i)).
\end{aligned}
$$

The proof that these trapezoids fit nicely together to a tiling of a rectangle is done inductively. For the organization of the inductive argument, the following lemma is useful.

Lemma 4.9. *The orientation B of G with $(i,j) \in B$ iff $y(i) < y(j)$ is a bipolar orientation with source s and sink t.*

Proof. This follows from the maximum principle for discrete harmonic functions, i.e, from the fact that such a function assumes its maximum at boundary vertices. □

As with squarings, it may happen that trapezoids degenerate and have zero area. We say that a Markov chain p on G is generic if this does not happen. In particular, this implies that $p(i,j) + p(j,i) > 0$ for all edges (i,j) of G and $y(i) \neq y(j)$ for every pair i,j of adjacent vertices.

Theorem 4.10 (Kenyon '98). *Let G be planar with s and t on the outer face. If p is a generic Markov chain on G and $\mathcal{T}$ is the trapezoidal dissection associated with (G,p) by the above construction, then $G = G_{\mathcal{T}}$ and p can be recovered from $\mathcal{T}$.*

Some special cases are particularly interesting:

- If p is reversible, i.e., if $\pi(i)p(i,j) = \pi(j)p(j,i)$ for all edges, then all trapezoids in the dissection are rectangles. (This is the case of planar electrical networks with edges of varying resistance.)
- If $p(i,j) = \frac{1}{\deg(i)}$, then $\pi(i) = \frac{\deg(i)}{2m}$ and the aspect ratios $m(i,j) \sim \pi(i)p(i,j)$ of all rectangles are equal. Hence, after scaling the dissection, we obtain a squaring.

Kenyon [30] also considers dissections of more general trapezoidal shapes than rectangles. Markov chains that yield dissections into nontrapezoidal shapes are considered in [36].

5 Square Duals

In this section, we review a result of Schramm [49]. He proves the existence of rectangular duals where all rectangles are squares for a large class of triangulations. The approach is based on *extremal length*. To begin, we need some definitions. Throughout $G = (V,E)$ is a prescribed graph.

- Any function $m : V \to \mathrm{IR}^+$ is called a *discrete metric* on G.
- The length of a path γ in G is $\ell_m(\gamma) = \sum_{x\in\gamma} m(x)$.
- For $A,B \subset V$, let $\Gamma(A,B)$ be the set of A,B paths.
  The distance between A and B is defined as $\ell_m(A,B) = \min_{\gamma\in\Gamma(A,B)} \ell_m(\gamma)$.
- The *extremal length* of A,B is $L(A,B) = \sup_m \frac{\ell_m(A,B)^2}{||m||^2} = \sup_m \frac{\ell_m(A,B)^2}{\sum_v m(v)^2}$.
- A metric realizing the extremal length is an *extremal metric*.

Proposition 5.1. *For G and $A,B \subset V$, there is a (up to scaling) unique extremal metric.*

Proof. If m is extremal, then so is every positive scalar multiple of m. Therefore, we only have to look at metrics m with $\ell_m(A,B) = 1$. □

These metrics form a polyhedral set P described by the finitely many linear inequalities of the form $\sum_{x\in\gamma} m(x) \geq 1$ with $\gamma \in \Gamma(A,B)$.

The extremal metric is the unique $m \in P$ with minimal norm $||m|| = \sqrt{\sum_v m(v)^2}$. □

Proposition 5.2. *A squaring with A and B representing the top and bottom of the dissection induces an extremal metric on the rectangular dual graph G of the squaring.*

Proof. Let $h = \mathsf{height}(R)$ and $w = \mathsf{width}(R)$. We may assume $h \cdot w = 1$. Let $s : V \to \mathrm{IR}$ be the metric where $s(v)$ is the side length of the square of vertex v. Since $||s||^2 = \sum s(v)^2 = h \cdot w = 1$, we have $||s|| = 1$. □

Let m be any metric. For $t \in [0,w]$, the squaring induces a path γ_t consisting of the vertices v whose representing square is intersected by the vertical line $x = t$. By definition, $\ell_m(A,B) \leq \sum_{v\in\gamma_t} m(v)$.

$$w \cdot \ell_m(A,B) \leq \int_0^w \sum_{v\in\gamma_t} m(v)dt = \int_0^w \sum_{v\in V} m(v)\delta_{[v\in\gamma_t]}dt = \sum_{v\in V} m(v) \int_0^w \delta_{[v\in\gamma_t]}dt$$
$$= \sum_{v\in V} m(v)s(v) = \langle m,s\rangle \leq ||m|| \cdot ||s|| = ||m||$$

Hence,

$$\frac{\ell_m(A,B)^2}{||m||^2} \leq \frac{1}{w^2} = h^2 = \frac{h^2}{||s||^2}.$$

This shows that the metric s is extremal. □

The proposition, together with the uniqueness of extremal metrics, implies that for a given G, there can, up to scaling, only be a single square dual.

We next show the converse of the proposition: Extremal lengths can be used to get squarings. Let G be an inner triangulation of a 4-gon, i.e., a triangulation with one outer edge removed. Call the four vertices of the outer face s,a,t,b in counterclockwise order.

Now let m be an extremal metric with respect to a and b such that $||m|| = 1$ and define $h = l_m(a,b)$. For an inner vertex v of G, let $x(v)$ be the length of a shortest $s \to v$ path and $y(v)$ be the length of a shortest $a \to v$ path. The length of a path is, of course, taken with respect to m.

Proposition 5.3 (Schramm '93). *The squares* $Z_v = [x(v) - m(v), x(v)] \times [y(v) - m(v), y(v)]$ *for inner vertices of* G *together with four appropriate squares for the outer vertices yield a square contact representation of* G *in the rectangle* $R = [0, h^{-1}] \times [0, h]$. *Moreover, if* G *has no separating 3- and 4-cycles, then there are no degenerate squares; i.e.,* $m(v) > 0$ *for all* v.

Proof. The first claim is that all edges are represented as contacts or intersections. Let (u,v) be an interior edge of G and $u,v \neq t$. From $x(v) \leq x(u) + m(v)$ and $x(u) \leq x(v) + m(u)$ and the corresponding inequalities for $y(u)$ and $y(v)$, it follows that $Z_u \cap Z_v \neq \emptyset$. The same property for edges of the form (u,t) is nontrivial. Schramm [49] has a direct argument, while Lovász [38] argues with blocking polyhedra and shows that m is an extremal metric with respect to s and t. We skip this part of the proof. □

The next claim is that the squares cover R; i.e., $R \subseteq \bigcup_v Z_v$. Suppose that there is a point p in R that is not contained in a Z_v. For each v, choose a representative point $q_v \in Z_v$ and draw the edges of G such that an edge (u,v) is represented by a curve connecting q_u and q_v that stays in $Z_u \cup Z_v$. The simplest choice for the edge may be to represent it as union of straight segments $[q_u, q_{uv}]$ and $[q_{uv}, q_v]$ for some $q_{uv} \in Z_u \cap Z_v$. With triangle (u,v,w) in G, let $T_{(u,v,w)}$ be the topological triangle enclosed by the edges connecting q_u, q_v, and q_w in R. By following a generic ray starting at p and considering situations where the ray crosses a curve representing an edge, it can be verified that p is contained in an odd number of the triangles $T_{(u,v,w)}$. Therefore, there is at least one triangle $T_{(u,v,w)}$ covering p. Consider $S = Z_u \cup Z_v \cup Z_w$. The boundary of $T_{(u,v,w)}$ is a closed curve γ contained in S such that $p \notin S$ is in the interior of γ. Since this is impossible for a union S of three squares, we have a contradiction. This proves the claim.

Since $\mathsf{area}(R) = 1 = ||m||^2 = \sum_v m(v)^2 = \sum_v \mathsf{area}(Z_v)$, it follows that intersections of the squares are confined to their boundaries, and we have a tiling of R with squares. We also know that all edges of G are represented by contacts. With a counting argument, it can be shown that these have to be all contacts between the squares except for

- Additional corner-to-corner contacts at points where four squares meet.
- Contacts where at least one of the two participating squares is a degenerate square of size 0.

This shows that the extremal length yields a square contact representation of G.

Finally, we show that in the absence of separating 3- and 4-cycles, all squares have nonzero area. Let W be the set of vertices of a connected component of the subgraph of G induced by vertices with degenerate squares, i.e., with $m(v) = 0$. All the vertices of W are represented by the same point p in R. This point p must also be contained in all Z_v, where $v \notin W$ is a neighbor of some $w \in W$. These at most four vertices v form a cycle separating W from the outer vertices. □

5.1 Square Duals and Transversal Structures

We now propose an alternative method to approach the problem of finding a square dual for an inner triangulation of a 4-gon. The approach is similar to what we have done in Sect. 4.2. First, we encode a rectangular dual as a graph with additional structure, in this case a transversal structure. From the transversal structure, we extract a system of linear equations such that a nonnegative solution yields the metric information needed for the square dual. If the solution has negative variables, we do not know how to use the solution to get a square dual. In this respect, the situation is more complicated than in Sect. 4.2. However, the signs in the solution provide a rule for changing the transversal structure. With the new transversal structure, we can proceed as before. Unfortunately, we cannot yet prove that the procedure stops, i.e., that at some iteration the solution is nonnegative and we get the square dual we are looking for.

Let G be an inner triangulation of a 4-gon with outer vertices s, a, t, b in counterclockwise order. A *transversal structure* for G is an orientation and 2-colorings of the inner edges of G such that

1. All edges incident to s, a, t, and b are blue outgoing, red outgoing, blue ingoing, and red ingoing, respectively.
2. The edges incident to an inner vertex v come in clockwise order in four nonempty blocks consisting solely of red ingoing, blue ingoing, red outgoing, and blue outgoing edges, respectively.

Transversal structures have been studied in [22, 23, 31]. The relevance of transversal structures in our context comes from the following simple proposition. See Fig. 23.

Proposition 5.4. *Transversal structures of an inner triangulation G of a 4-gon with outer vertices s, a, t, b are in bijection with combinatorially different rectangular dissections R with rectangular dual $G \setminus \{s, a, t, b\}$, where the rectangles of vertices*

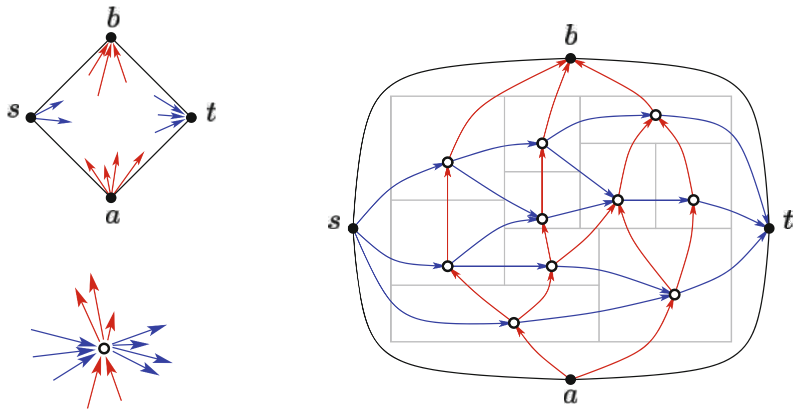


Fig. 23 The two local conditions and an example of a transversal structure together with a corresponding dissection

adjacent to s, a, t, and b touch the left, lower, right, and upper boundary of R, respectively.

Based on a transversal structure T, we want to write down a system of linear equations such that a nonnegative solution of the system yields a square dissection with T as the underlying transversal structure. For an inner vertex v, let x_v be a variable intended to represent the size of the square representing v. For a directed edge (u,v) in T, let x_{uv} be a variable representing the length of the contact between the rectangles representing u and v. The edges incident to a vertex v are partitioned into the four nonempty classes $\mathcal{R}^+(v)$, $\mathcal{B}^+(v)$, $\mathcal{R}^-(v)$, and $\mathcal{B}^-(v)$, where the letter indicates the color of the edge and the sign denotes whether the edge is outgoing or incoming. With vertex v, we associate four equations:

$$\sum_{(u,v)\in\mathcal{R}^+(v)} x_{uv} = x_v, \quad \sum_{(u,v)\in\mathcal{B}^+(v)} x_{uv} = x_v, \quad \sum_{(v,u)\in\mathcal{R}^-(v)} x_{vu} = x_v, \quad \sum_{(v,u)\in\mathcal{B}^-(v)} x_{vu} = x_v.$$

To forbid the trivial solution, we require that the width of the enclosing rectangle be 1. This is done by adding the equation $\sum_{(u,b)\in\mathcal{R}^-(b)} x_{ub} = 1$. Collecting the coefficients of the equations in a matrix A_T, we find that the vector x of lengths of a square dissection is a solution to the system

$$A_T \cdot x = \mathbf{e_1}. \tag{3}$$

Fact 1. *If the solution vector x is nonnegative, then there is a square dissection with the lengths given by x. If the solution vector x is positive, the transversal structure corresponding to the square dissection is T.*

Fact 2. *The matrix A_T is nondegenerate; hence, there is a unique solution x to the system.*

A proof of Fact 2 can be given along the lines of the proof of Theorem 4.2. The idea is to interpret A_T as the adjacency matrix of a planar bipartite graph H. Terms in the Leibniz expansion of $\det(A_T)$ correspond to perfect matchings in H. It can be shown that H admits a perfect matching. Moreover, the sign of all perfect matchings is equal because all inner faces of H have length 10, i.e., have residue 2 modulo 4.

To deal with the case where the solution vector x has negative entries, we need more insight into transversal structures. Recall from Sect. 2.3 the definition of the *trimmed angle graph* $\check{A}(G)$ associated with a plane graph G. The vertex set of this graph consists of the primal vertex set V together with the dual vertex set save the dual of the unbounded face, i.e., $V^* \setminus \{f_\infty\}$. Edges of $\check{A}(G)$ correspond to incidences between vertices and bounded faces of G or equivalently to the internal angles of G. Inner faces of $\check{A}(G)$ correspond to the inner edges of G.

Given a transversal structure T on G, we define an orientation of $\check{A}(G)$ as follows: Orient the edge $\{v, f\}$ as $v \to f$ when the two edges incident to f and v have different colors; otherwise, the edge is oriented $f \to v$. It can be verified that

- $\mathsf{out\text{-}deg}(v) = 4$ for all inner vertices of G, $\mathsf{out\text{-}deg}(v) = 0$ for the four vertices of the outer face, and $\mathsf{out\text{-}deg}(f) = 1$.

An orientation of $\check{A}(G)$ obeying the above rules for the out-degrees is called an α_4-*orientation* (note that up to changing the role of the color classes, α_4-orientations are identical to the α_{skel} orientations from Sect. 2.3).

Fact 3. *Let G be an inner triangulation of a 4-gon with outer vertices s,a,t,b. Transversal structures of G are in bijection with α_4-orientations of $\check{A}(G)$.*

Fact 4. *Let x be the solution to the system of equations corresponding to the transversal structure T, and let $E^-(x)$ be the set of edges whose value in x is negative. The boundary $\partial^-(x)$ of the union of all faces of $\check{A}(G)$ corresponding to edges in $E^-(x)$ decomposes into directed cycles with respect to the α_4-orientation corresponding to T.*

Reverting a directed cycle in an α_4-orientation yields another α_4-orientation. Hence, reverting all edges of $\partial^-(x)$ yields another α_4-orientation, which corresponds to a new transversal structure T' of G.
The approach for computing a square dual for G is this:

- Compute a transversal structure T of G and the matrix A_T.
- Compute a solution x_T of $A_T \cdot x = \mathbf{e_1}$.

If all entries of x_T are nonnegative, we are done; based on x_T, we can build the square dissection for G. If there are negative entries in x_T, we can use the α_4-orientation to transform T into another transversal structure T' and iterate. We conjecture that independent of the choice of T, the sequence $T \to T' \to T'' \to$ has a finite length; i.e., there is a k such that the solution $x_{T^{(k)}}$ of the system corresponding to $T^{(k)}$ is nonnegative.

There is strong experimental support for the truth of the conjecture.

Acknowledgements I thank Thomas Picchetti for his implementation of the squaring algorithm of Sect. 5.1 and Julia Rucker and Torsten Ueckerdt for helpful discussions and their continuing interest in the topic. My thanks also go to Janos Pach for encouraging me to write about this topic and to the Bernoulli Centre for its hospitality.

This work was partially supported by DFG Grant FE-340/7-2 and the EUROGIGA Project GraDR.

References

1. M.J. Alam, T. Biedl, S. Felsner, M. Kaufmann, S.G. Kobourov, T. Ueckerdt, Computing cartograms with optimal complexity. In *Proceedings of the 2012 symposuim on Computational Geometry* (SoCG'12). ACM, Chapel Hill, North Carolina, USA, pp. 21–30 (2012)
2. A.L. Buchsbaum, E.R. Gansner, C.M. Procopiuc, S. Venkatasubramanian, Rectangular layouts and contact graphs. ACM Trans. Algorithms **4**(1), 8–28 (2008)
3. I. Benjamini, L. Lovász, Harmonic and analytic functions on graphs. J. Geom. **76**, 3–15 (2003)
4. B. Bollobas, *Modern Graph Theory* (Springer-Verlag, New york, 2002)
5. J. Bhasker, S. Sahni, A linear algorithm to find a rectangular dual of a planar triangulated graph. Algorithmica **3**, 247–278 (1988)
6. G.R. Brightwell, E.R. Scheinerman, Representations of planar graphs. SIAM J. Discr. Math. **6**, 214–229 (1993)
7. R.L. Brooks, C.A.B. Smith, A.H. Stone, W.T. Tutte, The dissection of rectangles into squares. Duke Math. J. **7**, 312–340 (1940)
8. J. Chalopin, D. Gonçalves, Every planar graph is the intersection graph of segments in the plane, in *Proceedings of the 41st ACM Symposium on Theory of Computing (STOC)*, Bethesda, Maryland, USA, pp. 631–638 (2009)
9. M. Dehn, Über zerlegung von rechtecken in rechtecke. Math. Ann. **57**, 314–332 (1903)
10. G. Di Battista, P. Eades, R. Tamassia, I.G. Tollis, *Graph Drawing* (Prentice Hall, Englewood Cliffs, NJ, 1999)
11. H. de Fraysseix, P. Ossona de Mendez, On topological aspects of orientation. Discr. Math. **229**, 57–72 (2001)
12. H. de Fraysseix, P. Ossona de Mendez, J. Pach, A left-first search algorithm for planar graphs. Discr. Comput. Geom. **13**, 459–468 (1995)
13. H. de Fraysseix, P. Ossona de Mendez, P. Rosenstiehl, Bipolar orientations revisited. Discr. Appl. Math. **56**, 157–179 (1995)
14. P.G. Doyle, J. Snell, *Random Walks and Electric Networks* (Math. Assoc. of America, 1984). http://math.dartmouth.edu/doyle/docs/walks/walks.pdf. Accessed on Sept 25, 2012
15. M. Eiglsperger, S. Fekete, G.W. Klau, Orthogonal graph drawing, in *Drawing Graphs: Methods and Models*. LNCS, vol. 2025 (Tutorial) (Springer-Verlag, 2001), pp. 121–171
16. D. Eppstein, E. Mumford, B. Speckmann, K. Verbeek, Area-universal rectangular layouts, in *Proceedings of the 25th Symposium on Computational Geometry*, Aarhus, Denmark, pp. 267–276 (2009)
17. S. Felsner, *Geometric Graphs and Arrangements*. Advanced Lectures in Mathematics (Vieweg Verlag Wiesbaden, 2004)
18. S. Felsner, Lattice structures from planar graphs. Electr. J. Comb. **11**(R15), 24 pp. (2004)
19. S. Felsner, É. Fusy, M. Noy, D. Orden, Bijections for Baxter families and related objects. J. Comb. Theor. A **18**, 993–1020 (2011)
20. S. Felsner, C. Huemer, S. Kappes, D. Orden, Binary labelings for plane quadrangulations and their relatives. Discr. Math. Theor. Comput. Sci. **12**(3), 115–138 (2010)
21. J.-H. Fan, C.-C. Lin, H.-I. Lu, H.-C. Yen, Width-optimal visibility representations of plane graphs, in *Proceedings of the 18th International Conference on Algebraic and Computation (ISAAC)*. LNCS, vol. 4835 (Springer, 2007), pp. 160–171

22. E. Fusy, Combinatoire des cartes planaires et applications algorithmiques. Ph.D. thesis, LIX Ecole Polytechnique, 2007. http://www.lix.polytechnique.fr/fusy/Articles/these_eric_fusy.pdf. Accessed on Sept 25, 2012
23. E. Fusy, Transversal structures on triangulations: a combinatorial study and straight-line drawings. Discr. Math. **309**(7), 1870–1894 (2009)
24. J. Graver, B. Servatius, H. Servatius, *Combinatorial Rigidity*. Grad. Stud. in Math., vol. 2 (American Math. Soc., Providence, RI, 1993)
25. X. He, On finding the rectangular duals of planar triangular graphs. SIAM J. Comput. **22**(6), 1218–1226 (1993)
26. P. Hliněný, Contact graphs of line segments are NP-complete. Discr. Math. **235**, 95–106 (2001)
27. I.-A. Hartman, I. Newman, R. Ziv, On grid intersection graphs. Discr. Math. **87**, 41–52 (1991)
28. R. Haas, D. Orden, G. Rote, F. Santos, B. Servatius, H. Servatius, D. Souvaine, I. Streinu, W. Whiteley, Planar minimally rigid graphs and pseudo-triangulations. Comput. Geom. **31**, 31–61 (2005)
29. X. He, H. Zhang, Nearly optimal visibility representations of plane graphs. SIAM J. Discr. Math. **22**, 1364–1380 (2008)
30. R. Kenyon, Tilings and discrete Dirichlet problems. Isr. J. Math. **105**, 61–84 (1998)
31. G. Kant, X. He, Regular edge labeling of 4-connected plane graphs and its applications in graph drawing problems. Theor. Comput. Sci. **172**, 175–193 (1997)
32. G. Kirchhoff, Über die Auflösung der lichungen, auf welche man bei der Untersuchung der linearen Verteilung galvanischer Ströme geführt wird. Ann. Phys. Chem. **72**, 497–508 (1847)
33. K. Koźmiński, E. Kinnen, Rectangular duals of planar graphs. Networks **15**, 145–157 (1985)
34. J. Kratochvíl, A. Kuběna, On intersection representations of co-planar graphs. Discr. Math. **178**, 251–255 (1998)
35. P. Koebe, Kontaktprobleme der konformen Abbildung. Ber. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Kl. **88**, 141–164 (1936)
36. R.W. Kenyon, S. Sheffield, Dimers, tilings and trees. J. Comb. Theor. Ser. B **92**, 295–317 (2004)
37. C.-C. Liao, H.-I. Lu, H.-C. Yen, Compact floor-planning via orderly spanning trees. J. Algorithms **48**(2), 441–451 (2003)
38. L. Lovász, Geometric representations of graphs. http://www.cs.elte.hu/lovasz/geomrep.pdf. Draft version 11 Dec 2009
39. A. Lee, I. Streinu, Pebble game algorithms and sparse graphs. Discr. Math. **308**, 1425–1437 (2008)
40. K. Miura, H. Haga, T. Nishizeki, Inner rectangular drawings of plane graphs. Int. J. Comput. Geom. Appl. **16**(2–3), 249–270 (2006) [in English]
41. G.L. Miller, J. Naor, Flow in planar graphs with multiple sources and sinks. SIAM J. Comput. **24**, 1002–1017 (1995)
42. S. Mozes, C. Wulff-Nilsen, Shortest paths in planar graphs with real lengths in $\mathcal{O}(n \log^2 n / \log\log n)$ time, in *ESA 2010*. LNCS, vol. 6347 (Springer, 2010), pp. 206–217
43. T. Nishiseki, M.S. Rhaman, *Planar Graph Drawing* (World Scientific, Singapore, 2004)
44. M. Rahman, T. Nishizeki, S. Ghosh, Rectangular drawings of planar graphs. J. Algorithms **50**(1), 62–78 (2004)
45. M. Rahman, S.-i. Nakano, T. Nishizeki, Box-rectangular drawings of plane graphs. J. Algorithms **37**(2), 363–398 (2000)
46. P. Rosenstiehl, Embedding in the plane with orientation constraints: the angle graph. Ann. NY Acad. Sci. **555**, 340–346 (1989)
47. P. Rosenstiehl, R.E. Tarjan, Rectilinear planar layouts and bipolar orientations of planar graphs, Discr. Comput. Geom. **1**, 343–353 (1986)
48. H. Sachs, Coin graphs, polyhedra, and conformal mapping, Discr. Math. **134**(1–3), 133–138 (1994)
49. O. Schramm, Square tilings with prescribed combinatorics. Isr. J. Math. **84**, 97–118 (1993)
50. J. Skinner, C. Smith, W. Tutte, On the dissection of rectangles into right-angled isosceles triangles, J. Comb. Theor. Ser. B **80**, 277–319 (2000)

51. K. Stephenson, *Introduction to Circle Packing: The Theory of Discrete Analytic Functions* (Cambridge University Press, Cambridge, 2005)
52. S. Sadasivam, H. Zhang, NP-completeness of st-orientations for plane graphs. Theor. Comput. Sci. **411**, 995–1003 (2010)
53. R. Tamassia, On embedding a graph in the grid with the minimum number of bends. SIAM J. Comput. **16**, 421–444 (1987)
54. C. Thomassen, Plane representations of graphs, in *Progress in Graph Theory*, ed. by Bondy, Murty (Academic Press, 1984), pp. 336–342
55. R. Tamassia, I.G. Tollis, A unified approach to visibility representations of planar graphs. Discr. Comput. Geom. **1**, 321–341 (1986)
56. W. Tutte, The dissection of equilateral triangles into equilateral triangles. Proc. Camb. Phil. Soc. **44**, 463–482 (1948)
57. W.T. Tutte, Squaring the square, in *2nd Scientific American Book of Mathematical Puzzles and Diversions*, ed. by M. Gardner (University of Chicago Press, Chicago, 1987). http://www.squaring.net/history_theory/brooks_smith_stone_tutte_II.html. Accessed on Sept 25, 2012
58. T. Ueckerdt, Geometric representations of graphs with low polygonal complexity. Ph.D. thesis, Technische Universität Berlin, 2011
59. P. Ungar, On diagrams representing graphs. J. Lond. Math. Soc. **28**, 336–342 (1953)
60. H. Zhang, X. He, Visibility representation of plane graphs via canonical ordering tree. Inform. Process. Lett. **96**, 41–48 (2005)

Convex Obstacle Numbers of Outerplanar Graphs and Bipartite Permutation Graphs

Radoslav Fulek, Noushin Saeedi, and Deniz Sarıöz

Abstract The disjoint convex obstacle number of a graph G is the smallest number h such that there is a set of h pairwise disjoint convex polygons (obstacles) and a set of n points in the plane [corresponding to $V(G)$]so that a vertex pair uv is an edge if and only if the corresponding segment $\overline{uv}$ does not meet any obstacle.

We show that the disjoint convex obstacle number of an outerplanar graph is always at most 5, and of a bipartite permutation graph at most 4. The former answers a question raised by Alpert, Koch, and Laison. We complement the upper bound for outerplanar graphs with the lower bound of 4.

1 Introduction and Preliminaries

An *obstacle representation* of a graph G, as first defined by Alpert et al. [1], is a straight-line drawing of G, together with a set of polygonal obstacles such that two vertices of G are connected with an edge if and only if the line segment between the corresponding points does not meet any of the obstacles. As they did, we assume the points corresponding to the graph vertices together with the polygon vertices are in general position (no three on a line). An obstacle representation of G with h obstacles is called an *h-obstacle representation* of G. The *obstacle number* of

R. Fulek (✉)
École Polytechnique Fédérale de Lausanne, Station 8, 1015 Lausanne, Switzerland
e-mail: radoslav.fulek@epfl.ch

N. Saeedi
The University of British Columbia 2366 Main Mall, Vancouver, BC, Canada V6T 1Z4
e-mail: noushins@cs.ubc.ca

D. Sarıöz
The Graduate School and University Center of The City University of New York,
365 Fifth Ave., New York, NY 10016, USA
e-mail: sarioz@acm.org

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_13,

G is the smallest number of obstacles needed in an obstacle representation of G. If we restrict the polygonal obstacles to be convex, we call such a representation a *convex obstacle representation*. *Convex obstacle number* and *h-convex obstacle representation* are defined similarly.

If the convex obstacles are required to be pairwise disjoint, we call such a representation a *disjoint convex obstacle representation*, and we define *disjoint convex obstacle number* and *h-disjoint convex obstacle representation* similarly. Surely, the convex obstacle number of a graph is at most its disjoint convex obstacle number. We conjecture that there are graphs having convex obstacle number strictly less than their disjoint convex obstacle number, so we reason about these two parameters separately.

In [3], it was shown that for any fixed h, the number of graphs on n (labeled) vertices with obstacle number at most h is at most $2^{O(hn\log^2 n)}$. From this, it follows that every graph class with $2^{\omega(n\log^2 n)}$ members on n vertices (such as the class of all bipartite graphs) has an unbounded obstacle number. It was also shown therein that the number of unlabeled graphs on n vertices with convex obstacle number at most h is at most $2^{O(hn\log n)}$. Since the number of planar graphs on n vertices is $2^{\Theta(n\log n)}$ (see [2] for exact asymptotics), the bounds given by [3] are inconclusive regarding the obstacle number or convex obstacle number of the class of planar graphs or a subclass.

Nonetheless, it was shown by Alpert et al. [1] that every outerplanar graph admits a 1-obstacle representation in which the obstacle is in the unbounded face. The same paper raised the question of whether the convex obstacle number of an outerplanar graph can be arbitrarily large. We answer this question in the negative. In particular, we prove the following two results regarding outerplanar graphs in Sects. 2 and 3, respectively.

Theorem 1. *The convex (and disjoint convex) obstacle number of every outerplanar graph is at most five.*

Theorem 2. *There are trees having disjoint convex obstacle number at least four.*

In Sect. 4, we prove the following regarding bipartite permutation graphs.

Theorem 3. *The convex (and disjoint convex) obstacle number of every bipartite permutation graph is at most four.*

2 Upper Bound on Convex Obstacle Number of Outerplanar Graphs

Proof of Theorem 1. We shall show that the convex obstacle number of every outerplanar graph is at most *five*, by giving a method to generate five convex obstacles that can represent any outerplanar graph. For a given connected outerplanar graph G, we first construct a digraph $\overrightarrow{G'}$ with certain properties, whose underlying

Input: A connected graph $G = G(V,E)$
. **Output**: The digraph $\overrightarrow{G}'$ called the BFS-digraph of G
.
.
$V' := V_0 :=$ singleton set with an arbitrarily chosen vertex of G (the BFS *root*)
$\overrightarrow{E'} := \emptyset$
$i := 0$
while $V' \neq V$ **do**
$V_{i+1} := \{v \mid u \in V_i, (u,v) \in E\} \setminus V'$
$V' := V' \cup V_{i+1}$
$\overrightarrow{E'} := \overrightarrow{E'} \cup \{\overrightarrow{(u,v)} \mid u \in V_i, v \in V_{i+1}, (u,v) \in E\}$
$i := i+1$
end while
return $\overrightarrow{G}'(V, \overrightarrow{E'})$

Algorithm 1: Algorithm to compute a BFS-digraph of a connected graph

graph is a subgraph of G. We call $\overrightarrow{G'}$ the BFS-digraph of G. We show an obstacle representation using *five* convex obstacles for the BFS-digraph, and then modify the representation without changing the number of obstacles to represent the graph G. We finally discuss how to accommodate the disconnected case, still with *five* obstacles.

2.1 Constructing the BFS-Digraph and Its Properties

Let G be a connected outerplanar graph. Perform the breadth-first search based Algorithm 1 on G that outputs a digraph, which we call the BFS-digraph of G, and denote by $\overrightarrow{G'}$. We say that a vertex of a BFS-digraph has depth i if its distance from the BFS root is i (Fig. 1).

Lemma 2.1. *A BFS-digraph $\overrightarrow{G'}$ of a connected outerplanar graph G has a straight-line drawing such that*

1. *Each vertex at depth i lies on the line $y = -i$.*
2. *Two edges are disjoint except possibly at their endpoints.*
3. *A vertical downward ray starting at a vertex v meets the graph only at v.*

Proof. Let $\overrightarrow{G'_i}$ denote the subgraph of $\overrightarrow{G'}$ induced on vertices at depth less than or equal to i. We show the existence of such a drawing by constructing it. We will proceed by induction on $\overrightarrow{G'_i}$.

Consider a planar embedding of the outerplanar graph G in which every vertex meets the outer face, with all vertices on a circle having the *root* as its topmost point. From now on, we do not distinguish between a graph and its embedding. Draw the

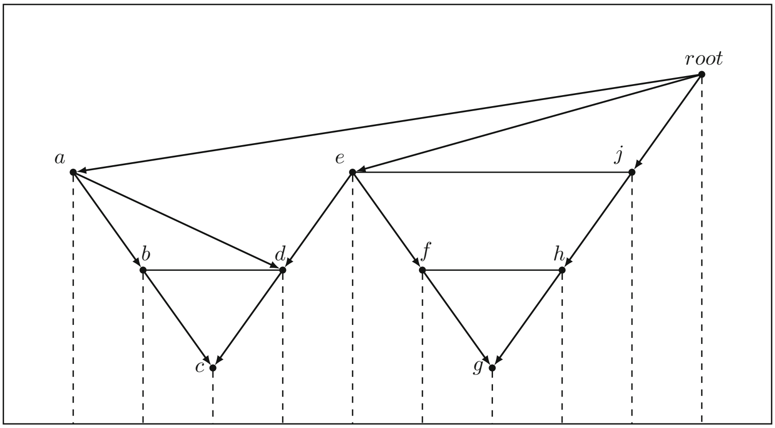


Fig. 1 A BFS-digraph of an outerplanar graph G drawn to exhibit the three properties in Lemma 2.1. The edges without arrows correspond to edges of G that are not in the digraph. For a given outerplanar graph G, regardless of the choice of the BFS root, there is a drawing of the resulting BFS-digraph that satisfies the three properties and induces a straight-line outerplanar drawing of G

root on the line $y = 0$. Then draw all its neighbors on the line $y = -1$ and to its left, preserving their order in G. All arcs corresponding to the edges between the root and its neighbors are oriented downward. So far we have $\overrightarrow{G'_1}$, which satisfies the desired properties. We now show how to extend for $i \geq 1$ an embedding of $\overrightarrow{G'_i}$ with the desired properties to an embedding of $\overrightarrow{G'_{i+1}}$ with the desired properties. Let $v_{i,1}, v_{i,2}, \ldots, v_{i,\ell} = root$ denote the vertices of $\overrightarrow{G'_i}$ in left-to-right order. For the sake of brevity, let $v_{i,0}$ also denote the *root*. The depth $i+1$ neighbors of a vertex $v_{i,k}$ in V_i lie in G either on the clockwise arc from $v_{i,k}$ to $v_{i,k-1}$ or on the counterclockwise arc from $v_{i,k}$ to $v_{i,k+1}$. Otherwise, G is not planar, or its vertices are not in convex position. We refer to the depth $i+1$ neighbors of $v_{i,k}$ on the clockwise arc from $v_{i,k}$ to $v_{i,k-1}$ as the *left* children of $v_{i,k}$, and those on the counterclockwise arc from $v_{i,k}$ to $v_{i,k+1}$ as the *right* children of $v_{i,k}$. Note that for vertices $v_{i,j}$ and $v_{i,j+1}$, the rightmost child of $v_{i,j}$ lies before or at the same place as the leftmost child of $v_{i,j+1}$. We apply the following steps for each $v_{i,k}$ in V_i:

- Put the left children of $v_{i,k}$, in order of clockwise proximity in G to $v_{i,k}$, on the line $y = -(i+1)$ so that they are to the left of $v_{i,k}$ and (unless $k = 1$) to the right of $v_{i,k-1}$.
- Put the right children of $v_{i,k}$, in order of counterclockwise proximity in G to $v_{i,k}$, on the line $y = -(i+1)$ so that they are between $v_{i,k}$ and $v_{i,k+1}$.
- Make sure that for every pair of vertices $v_{i,j}$ and $v_{i,j+1}$ in V_i, the rightmost child of $v_{i,j}$ and the leftmost child of $v_{i,j+1}$ preserve their order in G and are embedded once if they are one and the same.

Note that due to the outerplanarity of G, a right descendent and a left descendant of a vertex have no common descendants, rendering the last step possible. Therefore, the extended embedding represents $\overrightarrow{G'_{i+1}}$ and satisfies all three conditions. □

According to this embedding, we say that two vertices are *consecutive* if they are on the same horizontal line and there is no vertex between them.

Corollary 1. *A vertex has at most two parents. Moreover, if a vertex v has two parents, the parents are consecutive; and v is the rightmost child of its left parent, and the leftmost child of its right parent.*

Proof. If any of the conditions above does not hold, property 3 of Lemma 2.1 is violated. □

By Corollary 1, we also know two vertices at depth i have a common child only if they are consecutive.

Corollary 2. *Two consecutive vertices such that one is a left child and the other is a right child of the same parent do not have a common child.*

Proof. It directly follows from the third property of Lemma 2.1. □

2.2 5-Convex Obstacle Representation of the BFS-Digraph of a Connected Outerplanar Graph

We demonstrate a set of five convex obstacles and describe how to place vertices of $\overrightarrow{G'}$ to obtain a 5-convex obstacle representation for $\overrightarrow{G'}$. We first describe the arrangement of the set of obstacles. We have two disjoint convex arcs symmetric about a horizontal line, such that both arcs curve toward the line of symmetry. We consider the arcs to be parts of large circles, so that they behave like lines, except that they block visibilities among vertices put sufficiently near them. In the region bounded by the two arcs, we put three line obstacles, which form an S-shape with perpendicular joints, so that the S-shape is equally far from either arc, and the projection of the S-shape onto either arc covers the whole arc. We then disconnect the line obstacles by creating a small (and similar) aperture at each joint. The arrangement of the set of obstacles is shown in Fig. 2.

The key idea is to place all vertices of the graph sufficiently close to either of the arcs, and control the visibilities through the created apertures. For the sake of simplicity of exposition, from now on, we say a vertex is placed on an arc if it is sufficiently close to an arc. For each arc, the nearby and distant apertures are respectively called the *outgoing aperture* and the *incoming aperture*. For each vertex on an arc, we draw the outgoing edges through the outgoing aperture of its underlying arc. We parameterize the arcs such that the intersection points of the extended vertical line segment of the S-shape set at the arcs mark the zeros, and the positive axes of the lower arc and the upper arc point respectively to the right and to

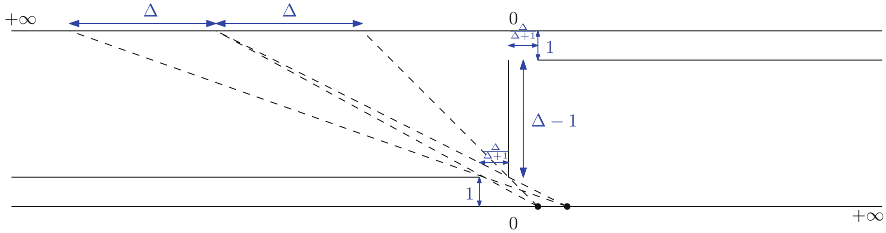


Fig. 2 The arrangement of the set of five convex obstacles

the left. Let $\Delta \geq 2$ be at least the maximum outdegree in $\overrightarrow{G'}$. We show if the S-shape is constructed so that

1. For two positive points unit distance apart on one arc, the parts of the opposite arc they see (through the outgoing aperture) share a single point.
2. Any point on an arc sees (through the outgoing aperture) an interval of length Δ of the other arc.

Then this obstacle set can represent BFS-digraphs of all connected outerplanar graphs.

We first investigate the structure of the set of obstacles to fulfill the conditions above. The distance between two compact subsets of the plane is the shortest distance between two of their respective points. Denote by w the aperture's width, denote by s the vertical segment's length (in the S-shape), and denote by x the distance between the S-shape and either arc.

Considering the arcs as lines, the first condition manifests if and only if $\frac{w}{1} = \frac{s+x}{s+2x}$, and the second condition holds if and only if $\frac{w}{\Delta} = \frac{x}{s+2x}$. These two equations require that $w = \frac{\Delta}{\Delta+1}$ and $s = (\Delta-1)x$, and we choose $x = 1$ to make things simple. We next show that the depicted set of obstacles represents any BFS-digraph. (Surely, the obstacle set depends on Δ, which is conditioned on $\overrightarrow{G'}$, and to list vertex coordinates of the polygonal obstacles we would also need to know the maximum depth in $\overrightarrow{G'}$, as we will discuss, so strictly speaking we have an obstacle set *template*.)

Proposition 1. *The arrangement of five convex obstacles shown in Fig. 2 represents BFS-digraphs of all connected outerplanar graphs.*

Proof. We give an algorithm to place the vertices of a connected BFS-digraph $\overrightarrow{G'}$ so that, together with the set of obstacles, they form an obstacle representation of $\overrightarrow{G'}$. We consider the two arcs in the obstacle set as lines; after all vertices are placed, we curve them a bit—just so that they block visibilities among vertices on them. This way, we ignore visibilities among vertices on the same arc (when considered as a line) and show that the set of obstacles represents $\overrightarrow{G'}$.

Consider a drawing of $\overrightarrow{G'}$ that satisfies the conditions in Lemma 2.1. From now on, by $\overrightarrow{G'}$ we refer to this embedding. Place the root of $\overrightarrow{G'}$ at coordinate 1 of the lower arc. We get a representation of $\overrightarrow{G'_0}$, where $\overrightarrow{G'_i}$ denotes the induced subgraph of $\overrightarrow{G'}$ containing all vertices at depth at most i. Suppose $\overrightarrow{G'_i}$ is represented such that

1. All vertices at an even depth are placed on the lower arc, and all vertices at an odd depth are placed on the upper arc.
2. On each arc, vertices at different depths are *well separated*; i.e., arc intervals containing all vertices at the same depth are disjoint.
3. Vertices of each depth preserve their ordering in $\overrightarrow{G'}$.
4. Every two consecutive vertices are at least one unit apart.

Note that by preserving the order, we mean if a vertex is to the left of some other vertex v in $\overrightarrow{G'}$, it gets a smaller coordinate than v when put on an arc.

Now, we describe how to add vertices at depth $i+1$ to obtain a representation of $\overrightarrow{G'_{i+1}}$ satisfying the conditions above. Let $v_{i,j}$ denote the jth vertex at depth i and let $[a_{i,j}, b_{i,j}]$ denote the interval of the opposite arc that is visible from $v_{i,j}$ through the outgoing aperture. For each vertex $v_{i,j}$ at depth i in the representation of $\overrightarrow{G'_i}$, we add its children on the opposite arc as follows:

- If $v_{i,j}$ has a common child with its immediate preceding vertex in V_i, put its leftmost child at $a_{i,j}$; otherwise, put the leftmost child at $a_{i,j} + \frac{1}{2}$.
- If $v_{i,j}$ has a common child with its immediate next vertex in V_i, put its rightmost child at $b_{i,j}$; otherwise, put the rightmost child at $b_{i,j} - \frac{1}{2}$.
- Put the remaining left children, preserving their ordering, after the leftmost one so that all left children are one unit apart.
- Put the remaining right children, preserving their ordering, before the rightmost one so that all right children are one unit apart.

Since every point sees an interval of length Δ, we know $b_{i,j} = a_{i,j} + \Delta$. Thus, as each vertex has at most Δ children, by performing the above algorithm, the rightmost left child is placed before the leftmost right child, and are at least one unit apart. Therefore, all consecutive pairs of vertices are of distance at least 1. Moreover, we know every two points, which are one unit apart, have a common point-of-sight; that is, the greatest point-of-sight of the smaller point equals the smallest point-of-sight of the greater one. By Corollaries 1 and 2, we know that if two vertices have a common child, then they are consecutive; and they are not right and left children of the same parent. Therefore, the presented algorithm put vertices so that two vertices at depths i and $i+1$ are visible in the representation if and only if they are connected in $\overrightarrow{G'}$. Conditions 1, 3, and 4 are surely satisfied after performing the algorithm. Since a vertex sees no other vertex except through the apertures, to complete the proof, what remains to be shown is that a vertex sees only its children through its outgoing aperture [and only its parent(s) through its incoming aperture]. To that end, next we prove that Condition 2 is satisfied, namely, that vertices at different depths are well separated: They lie in pairwise disjoint intervals.

Let I_0 denote the "interval" $[1,1]$ wherein the root is placed, and for every $i \geq 0$ let I_{i+1} denote the interval visible from I_i through the outgoing aperture. Since every vertex at depth i is in I_i, and I_i and I_{i+1} belong to different arcs, to prove Condition 2, it suffices to show that $I_i < I_{i+2}$ (i.e., every point in I_i has a smaller coordinate than every point in I_{i+2}) for every $i \geq 0$. If $I_i = [a,b]$, the structure of the obstacle set yields $I_{i+1} = [\Delta \times a, \Delta \times b + \Delta]$. Since $\Delta \geq 2$, this gives $I_0 < I_2$. By induction, we obtain that $I_i = [\Delta^i, 2\Delta^i + \sum_{j=1}^{i-1} \Delta^j]$ for every $i \geq 1$. Since $\Delta \geq 2$, for every $i \geq 1$, we have $2\Delta^i + \sum_{j=1}^{i-1} \Delta^j < 3\Delta^i < \Delta^{i+2}$; therefore, $I_i < I_{i+2}$.

Since we have previously ensured that a vertex v at depth i sees only its children through the outgoing aperture among all vertices at depth $i+1$, the well ordering of the intervals implies that v cannot see any other vertices through the outgoing aperture. By symmetry of sight, this implies that no vertex can see through its incoming aperture any vertex other than its parent(s).

This concludes the proof that we gave an obstacle representation of $\overrightarrow{G'}$. □

2.3 Adjusting the Representation for General Outerplanar Graphs

We first show how to modify the representation of a connected BFS-digraph $\overrightarrow{G'}$ to accommodate its corresponding outerplanar graph G. We know that the underlying graph of $\overrightarrow{G'}$ and G are the same, except that $\overrightarrow{G'}$ has no edge between two vertices at the same depth. Since G is an outerplanar graph, the extra edges of G, if any, are such that they connect two consecutive vertices. Therefore, to allow the existence of extra edges in the representation, we simply shave off the portion of the arc between their endpoints.

Now, we adapt this idea for disconnected outerplanar graphs. Let $C_1, C_2, \ldots, C_n$ be the components of a given outerplanar graph. Let $\overrightarrow{C_i'}$ be a BFS-digraph of C_i, as defined in Sect. 2.1. Let $\Delta \geq 2$ be at least the maximum outdegree among all BFS-digraphs, and construct the obstacle set template as before. Now, let L denote the maximum depth among all $\overrightarrow{C_i'}$. We declare I_0 to be the interval $[1,1]$ on one arc, and for every $i > 0$, we let I_i be the interval $[\Delta^i, 2\Delta^i + \sum_{j=1}^{i-1} \Delta^j]$ on the arc opposite interval I_{i-1}. The modified algorithm for representing a disconnected outerplanar graph is as follows. For each $\overrightarrow{C_i'}$, put its root at an arbitrary place in $I_{(i-1)(L+2)}$. Then carry out the algorithm described in Sect. 2.2 to place all vertices of $\overrightarrow{C_i'}$ for every i. This ensures that no vertex in C_i' can see a vertex of C_j' for any $i \neq j$. We then shave off the arcs as necessary to provide visibility among vertices at the same depth where desired.

We obtain a representation for an arbitrary outerplanar graph, concluding the proof of Theorem 1.

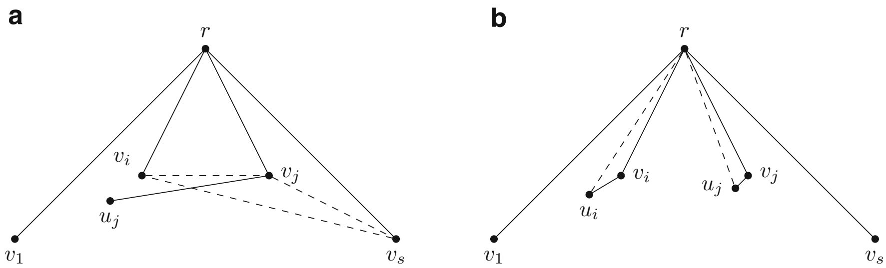


Fig. 3 For the proof of Lemma 3.2. Since all nonedges among $v_1, v_2, \ldots, v_s$ are blocked by a single convex obstacle O_1, these vertices are in convex position, and below r in the manner shown in both subfigures

3 Lower Bound on Disjoint Convex Obstacle Number of Outerplanar Graphs

For a rooted tree, we use the standard terminology—the depth of a vertex is its topological distance to the root, and the height of the tree is the maximum depth over all its vertices.

Proof of Theorem 2. Denote by $T_{k,h}$ the full complete k-ary tree with height h rooted at r. We will show that the disjoint convex obstacle number of $T_{k,3}$ is at least *four*, for k to be specified later. We say that two edges form a *crossing* if they meet at an internal point of both. (Recall that in an obstacle representation, no three vertices are collinear.) □

Lemma 3.2. *For every $m \in \mathbb{Z}^+$, there is a value of k such that $T_{k,2}$ has no m-convex obstacle representation without edge crossings.*

Proof. Denote by V_1 the set of vertices at depth 1, which is an independent set in $T_{k,2}$ of size k. For any given s, we can find a subset $V' \subseteq V_1$ of size s [provided large enough $k = k(s)$] such that every nonedge with both endpoints in V' is blocked by a common obstacle O_1. This is because we can assign every nonedge among V_1 to a single obstacle that blocks it to obtain an m-edge-coloring of a K_k induced on V_1, which by Ramsey's theorem has a monochromatic clique of size s for large enough k. The set V' lies in some half-plane having r on its boundary, without loss of generality, below a horizontal line; otherwise, r would be inside a triangle with vertices in V', yet no single convex obstacle could block all three sides of it without meeting an edge of $T_{k,2}$. Let us write $u < v$ whenever the triple ruv is counterclockwise. Let $v_1 < v_2 < \cdots < v_s$ denote the vertices in V'. For each $i : 1 < i < s$, let u_i denote a certain child of v_i. We claim that at least $(s-2)/2$ (not necessarily disjoint) convex obstacles are required to block the nonedges $ru_2, ru_3, \ldots, ru_{s-1}$.

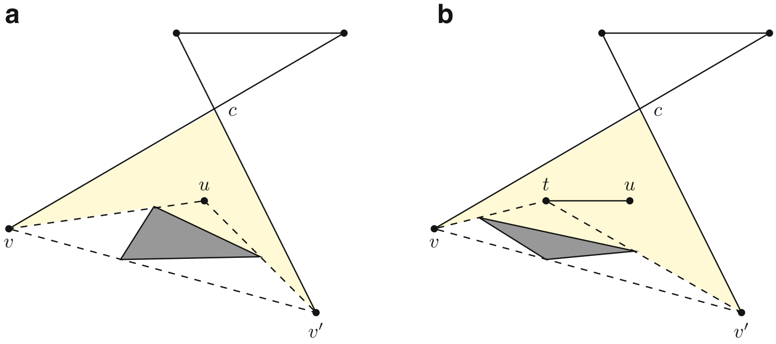


Fig. 4 Nonedges shown in each subfigure imply a respective minimal portion (*dark gray*) of an obstacle. The third edge of the path could have been incident on v or v', but this makes no difference. Only the obstacle that blocks vv' can be inside the convex angle $v'cv$

To prove the claim, assume for contradiction that some obstacle O' blocks three nonedges of the form ru_i. Then without loss of generality, for some pair $i < j$ such that ru_i and ru_j are blocked by O', both $u_i < v_i$ and $u_j < v_j$ hold. It must be that $v_i < u_j$; otherwise, $v_j u_j$ would cross an edge or meet O_1, which blocks both $v_i v_j$ and $v_j v_{j+1}$. See Fig. 3a. Choose two points $p_i \in ru_i \cap O'$ and $p_j \in ru_j \cap O'$. Then the segment $p_i p_j$ must intersect the union of the edges rv_i and $v_i u_i$. See Fig. 3b. By the convexity of O', we have a contradiction.

Thus, for $s \geq 2m+4$, at least $m+1$ convex obstacles are required if no edges cross. □

Assume for contradiction that we have a representation of $T_{k,3}$ with *three* pairwise disjoint convex obstacles O_1, O_2, and O_3.

If the endpoints of a crossing induced only the two edges (that is, an "X"-type crossing), at least four convex obstacles would be needed to block the nonedges, since any convex set that intersects two nonedges must meet an edge. However, no more than three edges can be induced by four vertices without forcing a cycle. Therefore, the four endpoints of every crossing induce a path with three edges.

By Lemma 3.2, we know that for large enough k, there are crossings within each subtree of $T_{k,3}$ isomorphic to $T_{k,2}$. Pick three crossings c_1, c_2, and c_3 in $T_{k,3}$, each in a subtree rooted at a different neighbor of the root vertex. For each $i \in \{1,2,3\}$, denote by $u_i v_i$ and $w_i z_i$ the edges of the crossing c_i, with the corresponding induced path on four vertices $P_i = u_i v_i w_i z_i$.

Let's first consider the case where the convex hulls of two of these paths, say P_1 and P_2, meet. If this is the case, with no vertex of P_1 being inside the convex hull of P_2 or vice versa, then some edge of P_1 must intersect some edge of P_2, inducing an "X"-type crossing that requires four obstacles. Hence, without loss of generality, some vertex u of P_1 is in the convex hull of P_2. See Fig. 4. Let c be the point of intersection of the two edges of P_2. Then u is inside some triangle vcv', where $v, v' \in P_2$. If vv' is an edge, then vcv' induces a bounded face, so uv would

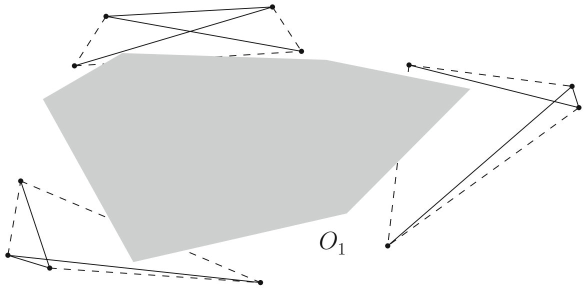


Fig. 5 Convex hulls of P_1, P_2, and P_3 are pairwise disjoint

require an obstacle in addition to the three required by P_2. Now, since u is inside a triangle vcv', the obstacle blocking vv' must also block uv and uv', but this forces all neighbors of u to be inside vcv'. Applying this argument to the neighbors of u in P_1 (recursively if needed), which satisfy the same conditions as u, we see that P_1 must be completely inside vcv'. But every nonedge of P_1 requires a distinct obstacle, at most one of which may coincide with one blocking vv' while none among them may coincide with any other obstacle, so five obstacles are required, a contradiction.

This means that the convex hulls of P_1, P_2, and P_3 are pairwise disjoint. Recall that for each of these paths P_i, each of the three nonedges of P_i must be blocked by a unique obstacle among three pairwise disjoint obstacles. Hence by the Jordan curve theorem, we get a contradiction (see Fig. 5).

4 Convex Obstacle Number of Bipartite Permutation Graphs

A *permutation graph* is a graph on $[n]$ according to a permutation $(\sigma_1, \sigma_2, \ldots, \sigma_n)$ of $[n]$ such that there is an edge between two elements $\sigma_i > \sigma_j$ whenever $i < j$. We show that the idea of having a small aperture between two classes of vertices, which are placed close to two convex obstacles, is readily extended to the class of bipartite permutation graphs.

Proof of Theorem 3. By a result from [4], a bipartite graph $G(V,E)$ is a permutation graph if and only if its two independent vertex classes V_1 and V_2 can be ordered such that the neighborhood of every vertex $u_i \in V_1$ forms an interval $[a_i, b_i]$ in V_2, and if $u_i < u_j$ for two vertices in V, then $a_i \leq a_j$ and $b_i \leq b_j$. □

We illustrate in Fig. 6 a set $\mathcal{C}$ of *four* disjoint convex obstacles allowing an obstacle representation of G. $\mathcal{C}$ consists of two convex arcs C_1 and C_2, and two vertical line segments (labeled $\mathcal{A}$) that form an aperture between C_1 and C_2.

Similarly to the treatment of the arcs in Sect. 2.2, we regard C_1 and C_2 as line segments, except that they block visibilities among graph vertices placed near them. For convenience, we shall speak of placing vertices of G *on* these arcs.

We put vertices of V_1 and V_2 on C_1 and C_2, respectively. Let $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ be the ordering of the vertices in V_1 and V_2 guaranteed by the

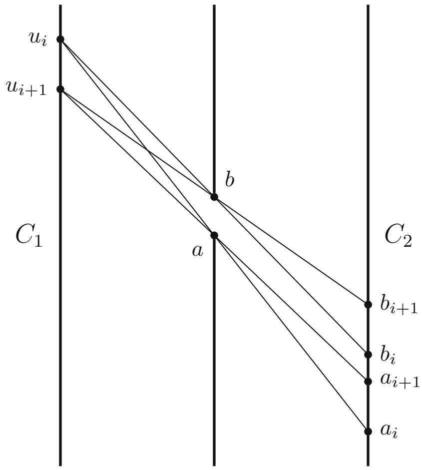


Fig. 6 The obstacles allowing a obstacle representation of a bipartite permutation graph

aforementioned result in [4]. We place the vertices, in order, inductively. In the basis step, we place u_1 arbitrarily on the relative interior of C_1. Let a_i and b_i denote the endpoints of the segment of C_2 that u_i can see through the aperture (see Fig. 6). We place neighbors of u_1 in the relative interior of segment a_1b_1 on C_2 so that the order of their y-coordinates corresponds to their order in V_2.

At an inductive step $i+1$, where $i \geq 1$, we place the $(i+1)$th vertex of V_1 together with its children, on the corresponding arcs as follows. We first find a consistent place for a_{i+1}. If the first neighbor w of the $(i+1)$th vertex (with regards to the order in V_2) is already placed on C_2, we pick a_{i+1} so that it precedes w (with regards to y-coordinate) and succeeds a_i and any other point already placed on C_2. Otherwise, we pick a_{i+1} so that it succeeds b_i. We place u_{i+1} at the intersection of C_1 and the line through a_{i+1} and a (see Fig. 6). The line through u_{i+1} and b intersects C_2 at point b_{i+1}, which has a higher y-coordinate than b_i. Therefore, we can place neighbors of u_{i+1} that are not neighbors of u_i on the nonempty line segment b_ib_{i+1}. This concludes the proof of Theorem 3.

Acknowledgements The authors are indebted to Filip Morić for fruitful discussions and for an idea that led to the proof of our upper-bound construction for outerplanar graphs. The authors would also like to thank the anonymous referee and Boris Aronov for providing useful feedback.

Radoslav Fulek and Noushin Saeedi's research was supported by Swiss National Science Foundation Grant 200021-125287/1 and the Bernoulli Center at EPFL. Noushin Saeedi's research was performed during a visit to EPFL during which she was partially supported by EPFL. Deniz Sarıöz's research was supported by NSA Grant 47180-0001.

References

1. H. Alpert, C. Koch, J. Laison, Obstacle numbers of graphs. Discr. Comput. Geom. **44**, 223–244 (2010)
2. O. Giménez, M. Noy, The number of planar graphs and properties of random planar graphs, in *2005 International Conference on Analysis of Algorithms*, ed. by C. Martínez. Discrete Mathematics and Theoretical Computer Science Proceedings AD, Barcelona, Spain, pp. 147–156 (2005)
3. P. Mukkamala, J. Pach, D. Sarıöz, Graphs with large obstacle numbers, in *36th International Workshop on Graph Theoretic Concepts in Computer Science (WG'10)*, Zaros, Crete, June 2010. Lecture Notes in Computer Science (LNCS), vol. 6410 (Springer-Verlag Berlin, Heidelberg, 2010), pp. 292–303
4. J. Spinrad, A. Brandstädt, L. Stewart, Bipartite permutation graphs. Discr. Appl. Math. **18**(3), 279–292 (1987)

Hanani–Tutte, Monotone Drawings, and Level-Planarity*

Radoslav Fulek, Michael J. Pelsmajer, Marcus Schaefer, and Daniel Štefankovič

Abstract A drawing of a graph is *x-monotone* if every edge intersects every vertical line at most once and every vertical line contains at most one vertex. Pach and Tóth showed that if a graph has an x-monotone drawing in which every pair of edges crosses an even number of times, then the graph has an x-monotone embedding in which the x-coordinates of all vertices are unchanged. We give a new proof of this result and strengthen it by showing that the conclusion remains true even if adjacent edges are allowed to cross each other oddly. This answers a question posed by Pach and Tóth. We show that a further strengthening to a "removing even crossings" lemma is impossible by separating monotone versions of the crossing and the odd crossing number.

Our results extend to level-planarity, which is a well-studied generalization of x-monotonicity. We obtain a new and simple algorithm to test level-planarity in quadratic time, and we show that x-monotonicity of edges in the definition of level-planarity can be relaxed.

---

*An earlier version of this paper appeared in the proceedings of WG 2011.

R. Fulek
Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: radoslav.fulek@epfl.ch

M.J. Pelsmajer
Illinois Institute of Technology, Department of Applied Mathematics, Chicago, IL 60616, USA
e-mail: pelsmajer@iit.edu

M. Schaefer
Department of Computer Science, DePaul University, Chicago, IL 60604, USA
e-mail: mschaefer@cs.depaul.edu

D. Štefankovič (✉)
Computer Science Department, University of Rochester, Rochester, NY 14627-0226, USA
e-mail: stefanko@cs.rochester.edu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_14,

1 Introduction

The classic Hanani–Tutte theorem states that if a graph can be drawn in the plane so every pair of independent edges crosses an even number of times, then it is planar [3, 31]. (Two edges are *independent* if they do not share an endpoint.) There are many ways to look at this result: In algebraic topology it is seen as a special case of the van Kampen–Flores theorem [21, Chap. 5], which identifies obstructions to embeddability in topological spaces. This point of view leads to challenging open questions (see, for example, [22]); even in two dimensions—for surfaces—the problem is not understood well. (See [30] for a survey of what we do know.)

Here, we study a variant of the problem that was introduced by Pach and Tóth [24]. A curve is *x-monotone* if it intersects every vertical line at most once. A drawing of a graph is *x-monotone* if every edge is x-monotone and every vertical line contains at most one vertex. In this context, the natural analogue of the Hanani–Tutte theorem is that for any x-monotone drawing in which every pair of independent edges crosses an even number of times, there is an x-monotone *embedding* (that is, a crossing-free drawing) with the same vertex locations. The truth of this result was left as an open problem by Pach and Tóth. We prove this monotone Hanani–Tutte theorem as Theorem 3.1 in Sect. 3.

The weak version of the classic Hanani–Tutte theorem states that if a graph can be drawn so that *every* pair of edges crosses evenly, then it is planar. (For background and variants of the weak Hanani–Tutte theorem, see [30].) For x-monotone drawings, this translates to the claim that if there is an x-monotone drawing in which every pair of edges crosses an even number of times, then there is an x-monotone embedding with the same vertex locations. This weak monotone Hanani–Tutte theorem was first proved by Pach and Tóth.[1] We give a new proof of this result as Theorem 2.11 in Sect. 2.

Traditionally, Hanani–Tutte style results are obtained via characterizations by obstructions. This can lead to very slick proofs, like Kleitman's proof of the Hanani–Tutte theorem for the plane (using obstructions K_5 and $K_{3,3}$) [18]. There are two drawbacks to this approach: Complete obstruction sets are often unknown, e.g., for the torus or, in spite of several attempts (as discussed in [8]), for x-monotone embeddings. Secondly, this approach is of little help algorithmically. Pach and Tóth proved the weak monotone Hanani–Tutte theorem using an approach of Cairns and Nikolayevsky [2] with which they proved the weak Hanani–Tutte theorem for orientable surfaces. Our approach is based on a different line of attack that began in [27].

In Sect. 4 we establish a connection between x-monotonicity and another well-known graph drawing notion, level-planarity. Through this connection, the monotone Hanani–Tutte theorem (Theorem 3.1) leads to a simple, quadratic-time algorithm for recognizing level-planar graphs. While the best-known algorithm for this

[1] There is a gap in the original argument; an updated version is now available [24, 25].

problem runs in linear time [16], it is quite complicated. There have been previous claims for simple quadratic-time algorithms for level-planarity testing, which we discuss in Remark 4.3.

The condition that edges are x-monotone in the monotone Hanani–Tutte theorems can be replaced by a weaker notion we call x-bounded. Let $x(v)$ denote the x-coordinate of a vertex v located in the plane. An edge uv in a drawing is *x-bounded* if every interior point p of uv satisfies $x(u) < x(p) < x(v)$. That is, an edge is x-bounded if it lies strictly between its endpoints; it need not be x-monotone within those bounds (see Corollary 2.7 and Remark 3.2). As a consequence, we obtain a relaxed, yet equivalent, definition of level-planarity (Corollary 4.5). We also describe an even weaker condition (nearly x-bounded) in Sect. 2.1 and show that we can get similar results for it as well (see Lemma 2.8 and Remark 3.2).

The classic Hanani–Tutte theorems have extensions that bound crossing number in terms of odd crossing number and independent odd crossing number, with equality for very small values [23,27,28]. We will see in Sect. 5 that such extensions fail for monotone odd and monotone independent odd crossing numbers. Also, Theorem 2.11 may prompt the reader familiar with Hanani–Tutte style results (in particular, [23, Theorem 1] and [27, Theorem 2.1]) to ask whether a stronger result is true: a "removing even crossings" lemma, which would say that all even edges can be made crossing-free even in the presence of odd edges (while maintaining x-monotonicity and vertex locations). We will see in Sect. 5 that there cannot be any such lemma for x-monotone drawings.

We end this section by stating a few definitions. The *rotation* at a vertex is the clockwise ordering of edges at that vertex, in a drawing of a graph. The *rotation system* of a graph is the collection of rotations at its vertices. In an x-monotone drawing, the *right (left) rotation* at a vertex is the order of the edges leaving the vertex toward the right (left). Usually, we consider graphs, but we will also have cause to study *multigraphs*, which allow the possibility of having more than one edge between each pair of vertices. Our multigraphs will never have loops. For any graph G and $S \subseteq V(G)$, let $G[S]$ denote the *subgraph induced by S*, which is the graph on vertex set S with edge set $\{uv \in E(G) : u \in S, v \in S\}$.

2 Weak Hanani–Tutte for Monotone Drawings

An edge is *even* if it crosses every other edge an even number of times (possibly 0 times). A drawing is *even* if all its edges are even.

Theorem 2.11 (Weak Monotone Hanani–Tutte; Pach, Tóth [24, 25]). *If G has an x-monotone even drawing, then G has an x-monotone embedding with the same vertex locations and rotation system.*

Our goal in this section is to give a new proof of Theorem 2.11.

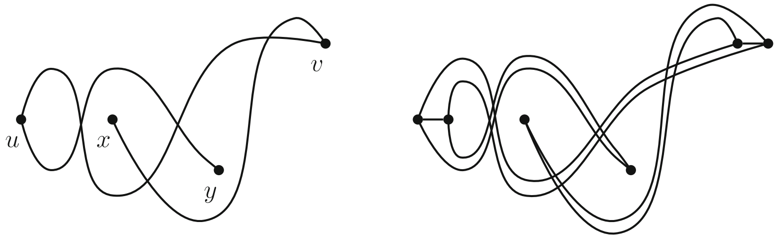


Fig. 1 *(Left)* An x-monotone even drawing. Since x is above uv and y is below uv, any equivalent x-monotone embedding with the same relative x-ordering of the vertices will have uv below x and above y. But then xv is above uy, so it is not equivalent. *(Right)* Essentially the same argument applies to this 2-connected example

Remark 2.2. By stretching and compressing and x-monotone drawing in the plane horizontally, we can change the x-coordinates of vertices arbitrarily as long as their relative x-order remains the same, and all the edges will remain x-monotone. We can also alter the y-coordinate of any vertex by stretching the plane vertically near that vertex, so that all edges remain x-monotone and all other vertices are fixed. Thus, we can modify an x-monotone drawing to relocate vertices arbitrarily, as long as the relative x-order of vertices is unchanged.

As a result, in any redrawing of an x-monotone drawing in which the relative x-order of the vertices does not change, we may as well assume that the location of every vertex has remained unchanged. Alternatively, we may instead assume that the vertices in an x-monotone drawing are located at the points $(1,0)$, ..., $(|V|,0)$. The same argument applies if the edges are x-bounded or nearly x-bounded (defined in Sect. 2.1) rather than x-monotone. (We briefly consider drawings with straight-line edges in Remark 2.4, and in that context the argument no longer works.)

The result claimed by Pach and Tóth in [24, Theorem 1.1] is almost the same as Theorem 2.11, but instead of maintaining rotations, they state that one can find an *equivalent* x-monotone embedding. Here, two drawings are equivalent if no edge changes whether it passes above or below a vertex. However, the example on the left in Fig. 1 shows that one cannot hope to maintain equivalence in this sense.

The proof in [24] contains a gap: It is not immediately clear how multiple faces that share a boundary can be embedded simultaneously.[2] Eliminating the gap requires dropping equivalence. Pach and Tóth have prepared an updated version of the paper that includes a more detailed argument [25].[3] As the graph on the right in Fig. 1 shows, the counterexample can be made 2-connected, so equivalence cannot be obtained by assuming 2-connectedness. On the other hand, see Corollary 2.5 for a positive result about equivalent redrawing.

[2] On page 42 of [24], in the text after Lemma 2.1, D_κ cannot necessarily be glued together without changing equivalence.

[3] In this newer version, equivalence is redefined to mean having the same rotation system.

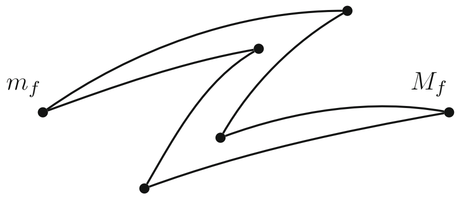


Fig. 2 Although we can draw the edge m_fM_f within the Z-shaped face, any subsequent x-monotone redrawing that maintains relative vertex x-order and rotation system will not be equivalent

In our proof of Theorem 2.11, we will repeatedly make use of a simple topological observation: Suppose we are given two curves (not necessarily monotone) that start at the line $x = x_1$ and end at the line $x = x_2$, and that lie entirely between $x = x_1$ and $x = x_2$. The two curves cross an even number of times if and only if they have the same vertical order at $x = x_1$ and $x = x_2$. (If they start or end in the same point p, the *vertical order at p* is determined by the vertical order in which they enter p.)

We will also use the following redrawing tool.

Lemma 2.3. *Let f be an inner face of an x-monotone embedding of a multigraph G, with m_f and M_f being the leftmost and the rightmost vertex of f. Add an edge m_fM_f to the embedding so that it lies in f. (Note that m_fM_f is not required to be x-monotone and and that there may be multiple ways of inserting m_fM_f into the rotations at m_f and M_f.) Then the resulting graph $G \cup \{m_fM_f\}$ has an x-monotone embedding with the same vertex locations and rotation system.*

Note that the redrawing in Lemma 2.3 destroys equivalence in the sense of Pach and Tóth [24]. This is necessarily the case; see Fig. 2 for an example.

Proof of Lemma 2.3. If G consists of multiple components, it is sufficient to prove the result for the component containing f and shift its embedding vertically so that it does not intersect any other component. This allows us to assume that G is connected. Then every face is bounded by a closed walk.[4] The boundary of f can be broken into two m_f, M_f-walks B_1, B_2, with B_1 starting above m_fM_f in the rotation at m_f, and B_2 starting below.

Let D_f be the drawing of G intersected with $U_f := \{(x,y) \in \mathbb{R}^2 : x(m_f) < x < x(M_f)\}$. ($D_f$ is a subset of the plane, not a graph.) We will locally redraw G in U_f so that m_fM_f can be inserted as a straight-line segment. For each (topologically) connected component Z of D_f, either (i) for every x between $x(m_f)$ and $x(M_f)$, there is a y-value of B_1 at x that is below all y-values of Z at x, or (ii) for any x between $x(m_f)$ and $x(M_f)$, there is a y-value of B_2 at x that is above all y-values of Z at x.

[4] *Walks* are like paths except that vertices and edges can be repeated. In a *closed walk*, the last vertex is the same as the first vertex.

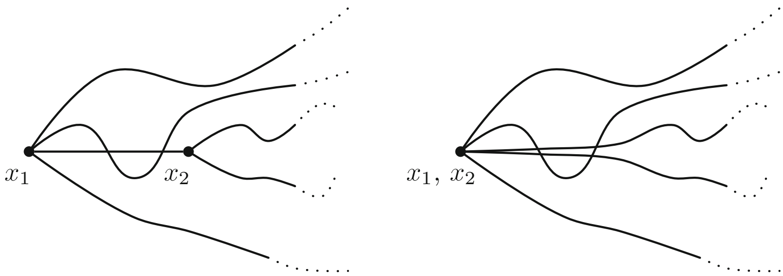


Fig. 3 How to contract edge x_2x_1 toward x_1 and merging rotations

Let Z_1 be the union of all components of the first type, and Z_2 be the union of all components of the second type. Let L be the line through m_f and M_f. We will show how to move Z_1 to the half-plane above L, without changing the x-value of any point in Z_1 while fixing the points on the boundary of U_f. Let P be an x-monotone curve with endpoints m_f and M_f that lies strictly below Z_1 in U_f (note that m_f and M_f do not belong to U_f). Now move every point v of Z_1 up by the vertical distance between P and L at $x = x(v)$. We proceed similarly to move Z_2 strictly below L, after which we can embed m_fM_f as L. The overall embedding is as desired. □

Proof of Theorem 2.1. We prove the following statement by induction on the number of vertices and then the number of edges:

If G is a multigraph that has an x-monotone drawing in which all edges are even, then G has an x-monotone embedding with the same vertex locations and rotation system.

In the base case, G consists of a single vertex, so the result is immediate. If G consists of multiple components, we can apply induction to each component and combine the drawings by stacking them vertically, that is, translate each component vertically so no two components intersect. Thus, we may assume that G is connected.

We first consider the case that there is more than one edge between the two leftmost vertices of G, $x_1 < x_2$. If there are several edges between x_1 and at x_2, say $e_1, \ldots, e_k$, these have to be consecutive in the rotations at both x_1 and x_2: This is trivial for the rotation at x_2, since all edges incident to it on the left have to go to x_1. Now suppose there is an edge $f = x_1x_\ell$, $\ell > 2$ so that f falls between two edges e_i and e_j in the rotation at x_1, $1 \leq i, j \leq k$. It is easy to see that f must cross either e_i or e_j oddly, which contradicts f being even, so such an edge does not exist. Hence, all edges between x_1 and x_2 are consecutive and, moreover, have mirror rotations at x_1 and x_2 (again a consequence of their being even). We can then replace them with a single edge e between x_1 and x_2. By induction, that reduced graph has the required embedding, and we can replace e with the multiple edges $e_1, \ldots, e_k$ obtaining the desired embedding of G.

Now, consider the case that there is only a single edge x_1x_2 between x_1 and x_2. We contract x_1x_2 by moving x_2 along the edge toward x_1 and inserting the right rotation of x_2 into the rotation at x_1 (see Fig. 3). Note that all edges remain even (since x_1x_2

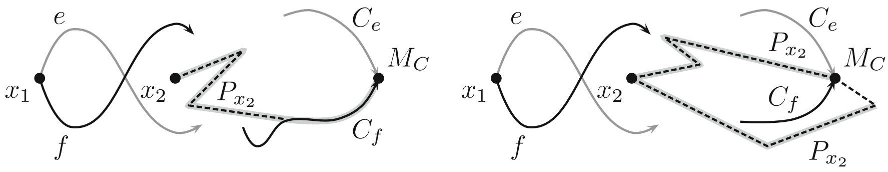


Fig. 4 *(Left)* Case (i): P_{x_2} is *dashed*, P'_{x_2} is the *thick gray* path from x_2 to M_C; *(right)* case (ii): both subcases are displayed: the top P_{x_2} stays to the left of M_C, while the bottom P_{x_2} passes to the right of M_C. P'_{x_2} as *thick gray* path in both cases

is even), so by induction the new graph has an x-monotone embedding in which $x_1 = x_2 < x_3 < \cdots < x_n$. We can now split the merged vertex into two vertices again and insert a crossing-free edge x_1x_2, obtaining an embedding of the original graph (since we kept the rotation) with the original rotation.

This leaves us with the case that there is no edge between x_1 and x_2. If $G - x_1$ consists of a single component, consider all edges $e_1, \dots, e_k$ leaving x_1. Each of these edges passes either above or below x_2. We claim that it is not possible that there are edges e and f so that e leaves x_1 above f but passes under x_2, while f passes above x_2: Assume for a contradiction that this is the case; pick a cycle C that contains both e and f (this cycle exists, since we assumed $G - x_1$ is a single component). Let M_C be the rightmost vertex of C. Consider the following two curves within C: C_e, which starts just below x_2 on e and leads to M_C, and C_f, which starts just above x_2 on f and leads to M_C. Note that since e leaves x_1 above f and C is a cycle consisting of even edges lying entirely between x_1 and M_C, curve C_e enters M_C above C_f. Pick a shortest path P_{x_2} from x_2 to C (such a path exists, since G is connected). We distinguish two cases (illustrated in Fig. 4).

(i) P_{x_2} *lies strictly to the left of* M_C. Without loss of generality, suppose that P_{x_2} ends on C_f. Let P'_{x_2} be the x_2, M_C-subpath of $P_{x_2} \cup C_f$. Since C_e and P'_{x_2} share no edges, and C_e passes below x_2, C_e must enter M_C below P'_{x_2} (all edges are even). However, the last part of P'_{x_2} belongs to C_f, so C_e enters M_C below C_f, which we know to be false.

(ii) P_{x_2} *contains a vertex at or to the right of* $x = M_C$. Let P'_{x_2} be the shortest subpath of P_{x_2} starting at x_2 and ending at or to the right of $x = M_C$. Since P_{x_2} has no edges in common with either C_e or C_f, P'_{x_2} enters M_C above C_e and below C_f if P'_{x_2} ends in M_C. Otherwise, P'_{x_2} passes M_C above C_e and below C_f. Since we know that C_e enters M_C above C_f, case (ii) also leads to a contradiction.

This establishes the claim that if e leaves x_1 above f, then it is not possible that e passes below x_2 while f passes above x_2. In other words, if some edge e starting at x_1 passes below x_2, then all edges starting below e at x_1 also pass below x_2. Hence, all edges passing below x_2 are consecutive at x_1 and so, perforce, are the edges passing above x_2. We can now add a new edge e from x_2 to x_1 that attaches in the rotation between the group of edges passing above x_2 and the edges passing below x_2.

This new edge will then be even, so we are in an earlier case that we know how to solve (by contracting the new edge, which reduces the number of vertices, then applying induction).

It remains to deal with the case that $G - x_1$ separates into multiple components. Let H_i', $i = 1,2$ be two of those components, and let H_i, $i = 1,2$ be H_i' together with its edges of attachment to x_1; that is, $H_i = G[\{x_1\} \cup V(H_i')]$. Note that the edges of H_1 and H_2 attaching to x_1 cannot interleave, meaning that at x_1 we cannot have edges e_1, e_2, f_1, f_2 in that order so that $e_i, f_i \in E(H_i)$ for $i = 1,2$. The reason is that e_i and f_i can be extended to a cycle $C_i \subset H_i$ and C_1 would cross C_2 an odd number of times if e_1, f_1 interleaves with e_2, f_2 at x_1. This implies that we can define a partial ordering $\prec$ on these components, where $H_1 \prec H_2$ if the edges (or edge) attaching H_1 to x_1 are surrounded (in the right rotation at x_1) by the edges attaching H_2 to x_1. Now let H be a minimal element of $\prec$; then the edges of H attaching to x_1 are consecutive at x_1. If H contains the rightmost vertex of G, then H is also a maximal element in $\prec$, so H cannot be the only minimal element of $\prec$; in this case, reassign H to another minimal element of $\prec$ that does not contain the rightmost vertex of G. Let $H' = H - \{x_1\}$.

Consider $G - V(H')$. By induction, there is an embedding of $G - V(H')$ that maintains the vertex locations and the rotation system. Let f be the face incident to x_1 into which H has to be reinserted (so that we recover the original rotation system). We can assume that f is not the outer face: If it is, we can make it an inner face by adding an edge from x_1 to the rightmost vertex of G. By Lemma 2.3, we can assume that the embedding has an x-monotone edge from x_1, starting where H' was attached in its rotation, to the rightmost vertex incident to f, which we call M_f. We can find an x-monotone embedding of H by induction. Note that all vertices of H must lie to the left of M_f, since otherwise an edge of H must have crossed an edge on the boundary of f oddly before $G - V(H')$ was redrawn using Lemma 2.3. But then we can insert the new embedding of H into the embedding of $(G - V(H')) \cup \{x_1 M_f\}$ near the edge $x_1 M_f$, such that there are no crossings, which gives us the desired embedding of G. □

Remark 2.4. There is a straight-line variant of Theorem 2.11 if we allow the y-coordinates of vertices to change. This has nothing to do with the Hanani–Tutte part of the result; it is entirely due to the fact that any x-monotone embedding can be turned into a straight-line embedding in which every vertex keeps its x-coordinate [7, 24]. This redrawing can lead to an exponential blow-up in the area required for the drawing [20].[5]

All redrawing steps in the proof of Theorem 2.11 maintain equivalence except for applications of Lemma 2.3. This part of the proof only arises if $G - \{x_1\}$ is not connected. Hence, if we can make an assumption on G so that this case never occurs, we can conclude that the resulting embedding is equivalent to the original

[5]The examples in that paper allow multiple vertices in each layer, but these can be replaced by the requirement that vertices are not too close to edges to which they are not incident.

drawing in the sense of Pach and Tóth [24]. We already saw that 2-connectedness is not sufficient; however, another notion is: A graph in which the vertices are ordered (from left to right, say) is a *hierarchy* if every vertex except the rightmost one has an edge leaving it toward the right [6].

Corollary 2.5. *If a hierarchy G has an x-monotone even drawing, then G has an equivalent x-monotone embedding with the same vertex locations and rotation system.*

Proof. This follows from the proof of Theorem 2.11. The only operation that changes equivalence of edges and vertices in that proof is the application of Lemma 2.3. If G is a hierarchy, $G-x_1$ consists of a single component, since any two vertices in $G-x_1$ are connected by a path (in a hierarchy, any two vertices must have a common ancestor). Since contracting the leftmost edge of a hierarchy results in another hierarchy, the result follows by induction. □

2.1 From x-Monotone to x-Bounded

The x-monotonicity assumption in Theorem 2.11 can be replaced by a weaker condition. Recall that an edge uv in a drawing is *x-bounded* if every interior point p of uv satisfies $x(u) < x(p) < x(v)$. That is, an edge is x-bounded if it lies strictly between its endpoints; it need not be x-monotone within those bounds.

Lemma 2.6. *Suppose we are given a drawing of a graph G with an x-bounded edge e. Then e can be redrawn, without changing the remainder of the drawing or the position of e in the rotations of its endpoints, so that e is x-monotone and the parity of crossing between e and any other edge of G has not changed.*

Proof. Suppose that $e = ab$ and let v be an arbitrary vertex between a and b: $x(a) < x(v) < x(b)$. Since e connects a to b, it has to cross the line $x = x(v)$ an odd number of times. Consequently, e crosses one of the two parts into which v splits $x = x(v)$ evenly, and e crosses the other part oddly. In a small neighborhood of $x = x(v)$, redraw G by pushing all crossings of e with $x = x(v)$ from the even side across v to the odd side (see top and middle parts of Fig. 5). Note that the odd side of $x = x(v)$ remains odd and there are no crossing with e left on the even side. Moreover, the parity of crossing between e and any other edge does not change since e is moved an even number of times across v. Repeat this for all v between a and b; now e only passes above or below each such v, never both. We can now deform e into an x-monotone edge connecting a and b, without having the edge pass over any vertices. Since the deformation does not pass over any vertex, it does not affect the parity of crossing between e and any other edge. This means we have found the redrawing required by the lemma (see middle and bottom parts of Fig. 5).

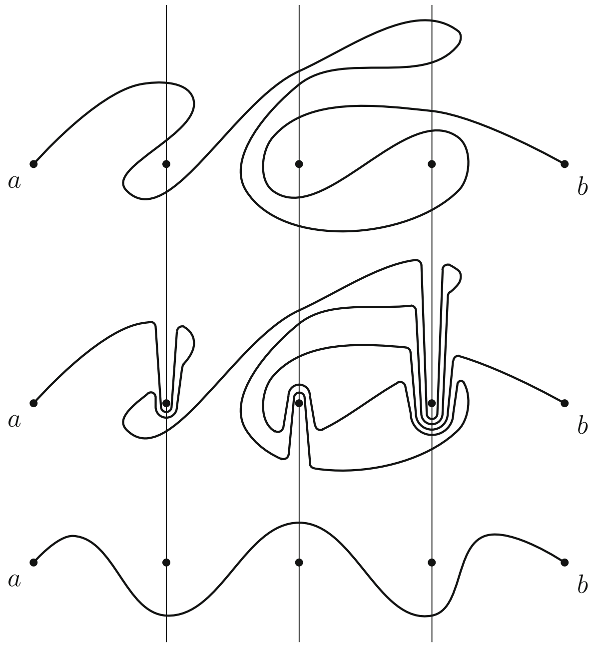


Fig. 5 How to redraw an x-bounded edge. (*Top*) Before the redrawing. (*Middle*) After pushing e off the odd parts. (*Bottom*) After deforming e into an x-monotone drawing

In hindsight, we see that the redrawing in Lemma 2.6 can be done quite efficiently: For each vertex v between a and b, we only need to know whether e passes oddly above or below it, and we can build a polygonal arc from a to b that passes each vertex on the odd side. □

Redrawing one edge at a time using Lemma 2.6 gives us the following strengthening of Theorem 2.11. Later, we will use that result to strengthen Theorem 3.1 and to show that x-monotone edges can be replaced by x-bounded edges in the definition of level-planarity (see Corollary 4.5 in Sect. 4.2).

Corollary 2.7. *If G has an even drawing in which every edge is x-bounded, then G has an x-monotone embedding with the same vertex locations and rotation system.*

Next, we give a condition weaker than x-bounded, for which we can prove some of the same results. Consider an edge e in a given drawing of a graph G with endpoints u, v such that $x(u) < x(v)$. Let C_e be the concatenation of e with the line segment from v to u. We say that e is *nearly x-bounded* if for every vertex z with $x(z) < x(u)$ or $x(z) > x(v)$, the winding number of C_e with respect to z is even.

Lemma 2.8. *Suppose we are given a drawing of a graph G with a nearly x-bounded edge e. Then e can be redrawn, without changing the remainder of the drawing, so that e is x-bounded and the parity of crossing does not change between e and any edge of G that is independent of e.*

Proof. We can gradually deform e to the line segment $\overline{uv}$, which causes e to become x-bounded. Suppose that e passed over the vertex z an odd number of times during the deformation. Since e is nearly x-bounded, it must be that $x(u) \le x(z) \le x(v)$. If $x(u) < x(z) < x(v)$, we can stretch e near the line $x = x(z)$ so that it passes over z once more, and e remains x-bounded. In the end, since e has passed over every vertex other than u and v an even number of times, the parity of crossing with e and any edge e' of G remains unchanged unless e' shares an endpoint with e. □

Although this does not allow us to directly generalize Corollary 2.7 to drawings with nearly x-bounded edges, we will apply Lemma 2.8 presently, in Remark 3.2.

3 Strong Hanani–Tutte for Monotone Drawings

Pach and Tóth [24] wrote, "It is an interesting open problem to decide whether [the conclusion of Theorem 2.11] remains true under the weaker assumption that any two *non-adjacent* edges cross an an even number of times." The goal of this section is to establish this result.

Theorem 3.1 (Monotone Hanani–Tutte). *If G has an x-monotone drawing in which every pair of independent edges crosses evenly, then G has an x-monotone embedding with the same vertex locations.*

Remark 3.2. Similarly to Theorem 2.11 and Corollary 2.7, the statement of Theorem 3.1 remains true if we only require edges to be x-bounded or nearly x-bounded rather than x-monotone: Simply redraw edges one at a time using Lemma 2.6 and/or Lemma 2.8, before applying Theorem 3.1.

In a proof of the standard Hanani–Tutte theorem, it is obvious that a minimal counterexample has to be 2-connected, since embedded subgraphs can be merged at a cut-vertex. Unfortunately, the merge requires a redrawing that does not maintain monotonicity, so here we must use structural properties that are more tailored to x-monotone redrawings. For a subgraph H of G, let $N(H)$ denote the set of neighbors of vertices of H in $G - V(H)$; that is, $N(H) := \{u : uv \in E(G), v \in V(H), u \in V(G) - V(H)\}$.

Lemma 3.3. *Suppose that G is a smallest (fewest vertices) counterexample to Theorem 3.1. Then*

(*i*) *G is connected.*
(*ii*) *G has no connected subgraph H and vertices $a, b \in V(G) - V(H)$ such that $x(a) < x(v) < x(b)$ for all $v \in V(H)$, $N(H) = \{a, b\}$, and $V(G) - (V(H) \cup \{a, b\}) \neq \emptyset$.*
(*iii*) *If G has a cut-vertex a and $G - \{a\}$ has a component H such that $x(a) < x(v)$ for all $v \in V(H)$, then H has only one vertex b, and G has no edge ac with $x(b) < x(c)$. Also, in this case G has no connected subgraph $H' \neq \emptyset$ so that $x(a) < x(v) < x(b)$ for all $v \in V(H')$, $a \in N(H') \neq \{a\}$, and $x(v) > x(b)$ for all $v \in N(H') - \{a\}$.*

Proof. If a smallest counterexample G is not connected, none of its components are counterexamples to Theorem 3.1. But then we could embed each component separately and stack the drawings vertically so they do not intersect each other, yielding an embedding of G. This contradiction establishes (i).

Consider (ii). Since G is a smallest counterexample, both $G - V(H)$ and $G[V(H) \cup \{a,b\}]$ have embeddings (both graphs are smaller than G by assumption). We can deform the crossing-free drawing of $G[V(H) \cup \{a,b\}]$ so that it becomes very flat. If $ab \in E(G)$, we can then insert this drawing into the drawing of $G - V(H)$ near the edge ab, without adding crossings. This gives us a crossing-free drawing of G, which is a contradiction. If $ab \notin E(G)$, then we add ab to the drawing of $G - V(H)$ so that it has no independent odd crossings (we will presently see how this can be done); the resulting $G - V(H) \cup \{ab\}$ has fewer vertices than G so it also has an embedding, and we can proceed as in the case that $ab \in E(G)$, removing the edge ab in the end.

When $ab \notin E(G)$, here is how we draw the edge ab with no independent odd crossings: Let P be any a,b-path with interior vertices in H. By suppressing the interior vertices of P, we can consider it an x-bounded edge (in the sense defined earlier) between a and b, so Lemma 2.6 tells us that we can draw an x-monotone edge that has the same parity of crossing with all edges of $G - V(H)$ as does P.

Finally, we consider (iii), where H is a component of $G - \{a\}$ so that $x(a) < x(v)$ for all $v \in V(H)$. Let b be the vertex with the largest x-value in H. If $|V(H)| > 1$, then we have case (ii) using $H := H - b$. Therefore, $|V(H)| = 1$ and $V(H) = \{b\}$. If G has an edge ac with $x(b) < x(c)$, we can first embed $G - \{b\}$ (since it is smaller than G), and then add ab and b to the embedding alongside of ac without crossings.

It remains to consider a connected subgraph $H' \neq \emptyset$ so that $x(a) < x(v) < x(b)$ for all $v \in V(H')$, $a \in N(H') \neq \{a\}$, and $x(v) > x(b)$ for all $v \in N(H') - \{a\}$. If there is an edge e not in H' with endpoints in H', we can replace H' by $H' \cup \{e\}$ and it still satisfies all the conditions; thus, we may assume that H' contains all such edges, i.e., that H' is an induced subgraph of G. By minimality, $G - \{b\}$ has an embedding. Of all the edges from a to H', let au be the one that is lowest in the rotation at a. Let f be the face in the drawing of G that lies immediately below au. Follow the boundary of f from a to u until it exits H' to a vertex c not in H'. If $c = a$, then H' could not have any neighbors v with $x(v) > x(b)$, a contradiction. The only other possibility is that $x(c) > x(b)$. Then by Lemma 2.3, we can add the edge ac to $G - \{b\}$ and obtain an embedding without introducing crossings. Since $x(a) < x(b) < x(c)$, we can instead add ab to the drawing without crossings, so G has an embedding, which is a contradiction. □

The proof of Theorem 3.1 now proceeds by induction on the number of *odd pairs* (pairs of edges that cross an odd number of times). Roughly speaking, if we encounter an odd pair (by necessity its edges are adjacent), we can either make it cross evenly or we are in a situation that has been excluded by Lemma 3.3. To realize this goal, we need additional intermediate results. These results are not about smallest counterexamples, but are true in general.

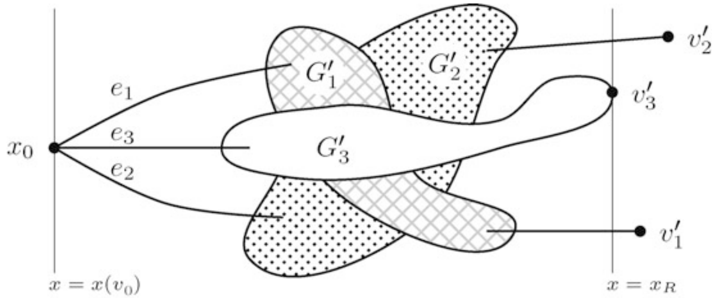


Fig. 6 Lemma 3.4

For the lemmas we introduce some new terminology generalizing our usual notion of lying above or below a curve to curves with self-intersections: Let C be a curve in the plane with endpoints p and r so that for every point $c \in C - \{p, r\}$, $x(p) < x(c) < x(r)$. (This is similar to the definition of an x-bounded edge except that we allow self-intersections.) Suppose that q is a point for which $x(p) \leq x(q) \leq x(r)$. Extend C via a horizontal ray from p to $x = -\infty$ and a horizontal ray from r to $x = \infty$, and consider the plane $\mathbb{R}^2$ minus that extended curve. We can 2-color its faces so that adjacent faces (faces whose boundaries intersect in a nontrivial curve) have opposite colors. We say that q is *above* (*below*) C if q lies in a face with the same color as the upper (lower) unbounded region.

In the following two lemmas, let G satisfy the assumption of Theorem 3.1; that is, we assume an even x-monotone drawing in which every pair of independent edges in G crosses evenly. Both lemmas deal with the following scenario: G contains three edges $e_i = v_0v_i$, $i \in \{1,2,3\}$ so that e_3 lies between e_1 and e_2 in the right rotation of v_0, with e_1 above e_2 at v_0, e_1 and e_2 cross oddly, and e_3 crosses each of the other two edges evenly.

Lemma 3.4. *For arbitrary $x_R > x(v_0)$, define G' as the graph induced by G on vertices v with $x(v_0) < x(v) \leq x_R$. Let G'_i be the component of G' that contains v_i. [If $x(v_i) > x_R$, then $G'_i = \emptyset$.]*

Suppose that G'_1, G'_2, G'_3 are pairwise disjoint and that for every i there is a path P_i (in G) from v_0 through e_i to some vertex v'_i satisfying $x(v'_i) \geq x_R$ so that all vertices v of P_i satisfy $x(v) \geq x(v_0)$. [If $G_i = \emptyset$, then let $E(P_i) = \{e_i\}$.] Then each G'_i has no neighbors (in G) to the left of $x(v_0)$, for $i \in \{1,2,3\}$ (Fig. 6).

Proof. By choosing each P_i to be minimal, we can assume that for every vertex v of P_i other than its final endpoint v'_i, we have $x(v_0) \leq x(v) < x_R$, and also $x(v'_i) \geq x_R$. Note that for any point in the plane q with $x(v_0) \leq x(q) \leq x_R$ that does not lie on the curve P_i, q is either *above* or *below* P_i in the sense defined just before Lemma 3.4.

Suppose, for a contradiction, that G'_i has a neighbor v' to the left of $x(v_0)$. Then we may let P'_i be a path from v_i to v' such that every vertex of $P'_i - v'$ is in G'_i. Fix j,k so that $\{i, j, k\} = \{1,2,3\}$.

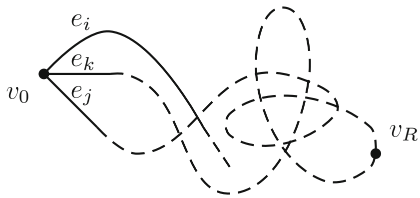


Fig. 7 e_i crosses e_j oddly and e_k evenly

The paths P_i' and P_j are disjoint, so every edge of P_i' crosses P_j evenly [as every pair $(e,f) \in (P_i', P_j)$ crosses evenly]. Every vertex of $P_i' - v'$ is between $x(v_0)$ and x_R, so if v_i is above P_j, then every vertex of $P_i' - v'$ is above P_j, and when the last edge of P_i' reaches the line $x = x(v_0)$, it must be above v_0. Likewise, if v_i is below P_j, then the last edge of P_i' must pass below v_0. Similarly, v_i is above (below) P_k if and only if the last edge of P_i' passes above (below) v_0.

We split the proof into cases. Suppose that $(i,j,k) = (1,2,3)$. Then e_i begins in the rotation at v_0 above e_j and e_k, and e_i crosses e_j oddly and e_k evenly. Since e_i crosses other edges of P_j and P_k evenly, v_i must be below P_j and above P_k. Then by the above/below argument in the previous paragraph, the last edge of P_i' must pass both below and above v_0, a contradiction. The case $(i,j,k) = (1,3,2)$ is the same, and the cases with $i = 2$ are symmetric.

Suppose that $i = 3$; without loss of generality, $(j,k) = (1,2)$. Then e_i begins in the rotation at v_0 below e_j and above e_k, and e_i crosses e_j and e_k evenly. Since e_i crosses other edges of P_j and P_k evenly, v_i must be below P_j and above P_k. Then by the earlier above/below argument, the last edge of P_i' must pass both below and above v_0, a contradiction. $\square$

Lemma 3.5. *Suppose that for some distinct $j,k \in \{1,2,3\}$, there is a cycle C that contains e_j and e_k such that every vertex v of C satisfies $x(v) \geq x(v_0)$. Let v_R be the vertex on C with the largest x-value. Let i be the unique index such that $\{i,j,k\} = \{1,2,3\}$. Suppose that v_i is not in C.*

Let G_i' be the component of $G - V(C)$ that contains v_i. Then every vertex v of G_i' satisfies $x(v_0) < x(v) < x(v_R)$.

Proof. Let P_j and P_k be the v_0, v_R-paths in C that contain e_j, e_k, respectively.

Suppose that $x(v_i) > x(v_R)$. First, consider the case $(i,j,k) = (1,2,3)$: Since e_j and e_k begin in the rotation at v_0 below e_i, and e_i crosses e_j oddly and e_k evenly, it must be that v_j is above e_i and v_k is below e_i (see Fig. 7). Every other edge of P_j crosses e_i evenly, so all its other vertices are also above e_i; likewise, every other vertex of P_k is below e_i. But then v_R lies above and below e_i, a contradiction. The case $(i,j,k) = (1,3,2)$ is the same, and the cases with $i = 2$ are symmetric. Suppose that $i = 3$; without loss of generality, $(j,k) = (1,2)$. Then in the rotation at v_0, e_j is above e_i and e_k is below e_i, and e_i crosses e_j and e_k evenly. Then v_j is above e_i and v_k is below e_i. Then (as seen earlier), every vertex of $P_j - v_0$ is above e_i and every vertex of $P_k - v_0$ is below e_i, a contradiction since v_R is in both.

Thus, we may assume that $x(v_0) < x(v_i) < x(v_R)$. As argued in the proof of Lemma 3.4, v_i is below P_j and above P_k, or v_i is above P_j and below P_k, depending on the order of values assigned to i, j, k.

Suppose that there is a path P_i' in $G - V(C)$ from v_i to a vertex v' with $x(v') < x(v_0)$ or $x(v') > x(v_R)$. Let P_i' be a minimal such path, so that it exits the region between lines $x = x(v_0)$, $x = x(v_R)$ with its last edge e'. P_i' is disjoint from P_j, so e' passes above (below) v_0 or v_R if and only if every vertex of P_i' is above (below) P_j. Likewise if we replace P_j by P_k. But then the vertices of P_i' are either all above P_j and P_k or they are all below P_j and P_k, including v_i, which contradicts what we have already shown for v_i. □

We are finally in a position to prove Theorem 3.1. We need one more piece of terminology: Consider two edges e and f that share the same right (or left) endpoint. The *distance* between e and f is the number of edge ends between the ends of e,f in the left (or right) rotation at their shared vertex. (We do not measure distance within the entire rotation; only within the right or left rotation.)

Proof of Theorem 3.1. Let G be a smallest (fewest vertices) counterexample to the theorem. By Lemma 3.3(i), G is connected. Fix an x-monotone drawing of G with the same vertex locations, which minimizes the number of odd pairs (that is, the number of pairs of edges crossing oddly). If there are no odd pairs, then Theorem 2.11 completes the proof.

Suppose that there are edges e_1 and e_2 that cross oddly. Then e_1 and e_2 have a shared endpoint v_0, and we may assume that v_0 is the left endpoint of e_1 and e_2. Choose e_1 and e_2 so that their ends at v_0 have minimum distance in the right rotation at v_0, with e_1 above e_2. Then e_1 and e_2 are not consecutive in the rotation at v_0; if they were, they could be redrawn so that they cross once more near v_0, by switching their order in the rotation at v_0; this contradicts the choice of drawing of G. So there is at least one edge incident to v_0 that lies between e_1 and e_2 in the rotation at v_0, and by minimality, all such edges cross each other evenly and cross both e_1 and e_2 evenly. Pick one such edge, e_3. Let v_1, v_2, v_3 be the right endpoints of e_1, e_2, e_3, respectively, and let G_0 be the subgraph of G induced by all vertices v fulfilling $x(v) \geq x(v_0)$.

Case 1. Vertices v_1, v_2, v_3 are in different components of $G_0 - v_0$.

In case 1, for each $i \in \{1,2,3\}$, consider the component of $G_0 - v_0$ that contains v_i and let v_i' be its vertex with largest x-value. Assign i, j, k so that $\{i, j, k\} = \{1,2,3\}$, and $x(v_i')$ is smaller than $x(v_j')$ and $x(v_k')$. Let $x_R = x(v_i')$ and apply Lemma 3.4, which defines G_i', G_j', G_k' and shows that G_i' has no neighbors to the left of $x(v_0)$. Then by Lemma 3.3(iii) (with $a = v_0$ and $H = G_i'$), G_i' has just one vertex $v_i = v_i'$ $(= b)$ and $x(v_i) > x(v_j)$ and $x(v_i) > x(v_k)$. Then G_j' and G_k' are nonempty, so they also have no neighbors to the left of $x(v_0)$. This contradicts the second part of Lemma 3.3(iii) with H' equal to G_j' (or G_k') restricted to vertices v with $x(v) \leq x_R$.

If we are not in case 1, let x_L be smallest such that the subgraph induced by $\{v \in V(G) : x(v_0) < x(v) \leq x_L\}$ has a component that contains at least two vertices of v_1, v_2, v_3. Then there is a cycle C that contains e_j and e_k for some distinct $k, j \in \{1,2,3\}$ and a vertex v_L such that $x(v_L) = x_L$ and $x(v_0) \leq x(v) \leq x(v_L)$ for all $v \in V(C)$. If $vv_L \in \{e_1, e_2, e_3\}$, then we may assume that C contains vv_L.

Let i be the unique index for which $\{i, j, k\} = \{1,2,3\}$. By the previous assumption, $v_i \neq v_L$. By Lemma 3.5, $x(v_i) < x(v_L)$ or $v_i \in V(C) - v_L$.

Suppose that there is a path Q from v_i to C so that $x(v_0) < x(v) < x(v_L)$ for all $v \in V(Q)$. Then $Q \cup e_i \cup C - v_L$ contains a cycle C' with e_i and either e_j or e_k. But every vertex v of C' satisfies $x(v_0) \leq x(v) < x(v_L)$ for all v in C', contradicting the choice of v_L.

It immediately follows that v_i is not in $V(C) - v_L$; also, $v_i \neq v_L$, so we may let G_i' be the component of $G - V(C)$ that contains v_i. By Lemma 3.5, G_i' lies between $x = x(v_0)$ and $x = x(v_L)$ (since $v_i \neq v_L$). The previous paragraph also implies that G_i' has no neighbors in $V(C) - \{v_0, v_L\}$. Let v_i' be the vertex of G_i' with largest x-value, let $x_R = x(v_i')$, and define G_i', G_j', G_k' according to Lemma 3.4 (and note that this doesn't alter G_i').

Case 2. G_i' is not adjacent to v_L.

(Same as case 1:) By Lemma 3.3(iii), G_i' has only the one vertex $v_i = v_i'$, and G_j' and G_k' are nonempty because $x(v_i)$ is greater than $x(v_j)$ and $x(v_k)$. Then we can apply Lemma 3.3(iii) with H' equal to G_j' (or G_k') restricted to the vertices with the x-coordinate smaller than $x(v_i')$, and we are done.

Case 3. There is an edge from G_i' to v_L.

Apply Lemma 3.3(ii) with $H = G_i'$. This completes the proof of the theorem. □

4 Level-Planarity Testing

The strong Hanani–Tutte theorem can be viewed as an algebraic characterization of planarity: Testing whether a graph is planar can be recast as solving a system of linear equations.[6] Unfortunately, the system has $|E| \cdot |V| = O(|V|^2)$ variables, which leads to an impractical $O(|V|^6)$ running time.[7]

Similarly, Theorem 3.1 can be viewed as an algebraic criterion for testing whether a graph has an x-monotone embedding, for a given x-coordinate order of the vertices. However, unlike the system of linear equations for planarity, the equations for x-monotonicity are so simple that solvability can be checked directly in quadratic

[6]Tutte presented his theorem as an algebraic characterization of planarity, but he did not investigate algorithmic implications [31]. Algebraic planarity testing based on the Hanani–Tutte characterization was probably first described by Wu [33,34] in a sequence of papers first published in Chinese in the 1970s.

[7]There are linear-time algorithms for planarity testing based on a Hanani–Tutte-like characterization, but they do not take the algebraic route [4,5].

time. We present the details of this algorithm in Sect. 4.1. In Sect. 4.2 we will see how to extend the algorithm to recognizing level-planar graphs, so we obtain a very simple, quadratic-time algorithm for level-planarity testing. Linear-time algorithms for this task are known but are quite complex. We discuss the rather confusing situation of algorithms for level-planarity testing in more detail in Remark 4.3.

4.1 Testing *x*-Monotonicity

How can we use Theorem 3.1 to test whether a given graph G with x-coordinates assigned to the vertices has an x-monotone embedding? Let D be an x-monotone embedding of G and let D' be an x-monotone drawing of G on the same vertex set. Pick any edge e in D' and continuously transform it into its drawing in D; we can assume that the edge remains x-monotone during the transformation. As the edge changes, its parity of intersection with any independent edge only changes when it passes over a vertex v (at which point its parity of intersection with every edge incident to v changes). The same effect can be achieved by making an (e,v)-*move*: Take e, and close to $x = x(v)$ deform it into a spike that passes around v. In other words, if G has an x-monotone embedding, then there is a set of (e,v)-moves that turns D' into a drawing in which every pair of independent edges crosses evenly. Since the reverse is also true, by Theorem 3.1, we now have an efficient test.

Theorem 4.1. *Given a graph G and a placement of the vertices of G in the plane, we can test in time $O(|V|^2)$ whether G has an x-monotone drawing on that vertex set.*

Proof. If G has an x-monotone embedding on the given vertex set, then no two vertices lie on a vertical line. As discussed in Remark 2.2, we can deform the plane so that the vertices are located at $(1,0)$, …, $(|V|,0)$, and the drawing will remain x-monotone—but it will remain an embedding as well. Thus, we can assume that the vertices are located at $(1,0)$, …, $(|V|,0)$.

Now draw each edge as a monotone arc above $y = 0$. Note that two edges cross oddly in this drawing if and only if their endpoints alternate in the order along the x-axis. By the discussion preceding the theorem, it is sufficient to decide whether there is a set of (e,v)-moves that turns this drawing into a drawing in which every pair of independent edges crosses evenly. We can model this using a system of equations: We introduce variables $x_{e,v}$ for each $e \in E$ and $v \in V$; $x_{e,v} = 1$ means an (e,v)-move is made, while $x_{e,v} = 0$ means it is not. For two edges $e = (e_1,e_2)$ and $f = (f_1,f_2)$ to intersect, their intervals on the x-axis have to overlap. And there are two cases: The endpoints alternate (and the edges cross oddly in the initial drawing) or they do not (and the edges cross evenly). Let's first consider the case $e_1 < f_1 < e_2 < f_2$. In the initial drawing, e and f cross oddly, so we must have $x_{e,f_1} = 1 - x_{f,e_2}$ for e and f to cross evenly. If $e_1 < f_1 < f_2 < e_2$, then e and f cross evenly, and we must have $x_{e,f_1} = x_{e,f_2}$ for e and f to cross evenly. Note that these equalities are the only conditions that affect whether e and f cross evenly. Hence, it is sufficient to set up this system of equations for all such pairs of edges e and f and solve it. This can be done using a simple depth-first search: Build a graph F on vertex set $E \times V$. Consider every pair of independent edges $e = (e_1, e_2)$ and $f = (f_1, f_2)$ in G, If $e_1 < f_1 < e_2 < f_2$, then add a red edge $((e, f_1), (f, e_2))$ to F. If $e_1 < f_1 < f_2 < e_2$, add a green edge $((e, f_1), (e, f_2))$ to F. Now perform a depth-first traversal of (the not necessarily connected) graph F. When starting the traversal at a new root, arbitrarily assign a value of 0 to the root variable. When following a green edge, assign the parent value to the child vertex, when following a red edge, swap 0 to 1, and vice versa. Whenever encountering a back edge, verify that the value assignment to the endpoints of the edge is consistent with its color (green for equal, red for different). If this test fails, the graph cannot be embedded. Otherwise, the depth-first search succeeds and the graph has an x-monotone embedding.

Since we can assume that G is planar, we know that $|E| \leq 3|V|$, and our algorithm runs in time $O(|V|^2)$ with a small constant factor. □

Remark 4.2. The $O(|V|^2)$ bound in Theorem 4.1 can be far from optimal since only (e, v)-moves for which v lies between the endpoints of e are possible. If we define the *layout complexity* of a graph with assigned x-coordinates as $|\{(e_1e_2, v) : x(e_1) < x(v) < x(e_2), v \in V(G), e_1e_2 \in E(G)\}|$, then the algorithm in Theorem 4.1 runs in linear time in the size of the layout. This measure seems fair if we actually want to draw the graph (since we have to know in what order edges pass a vertex).

4.2 Testing Level-Planarity

The definition of an x-monotone drawing does not allow two vertices to have the same x-coordinate. If we remove this restriction we enter the realm of leveled graphs: A *leveled graph* is a graph $G = (V, E)$ together with a *leveling* $\ell : V \to \mathbb{Z}$. A *leveled drawing* of (G, ℓ) is a drawing in which edges are x-monotone and $x(v) = \ell(v)$ for every $v \in V$. (G, ℓ) is *level-planar* if it has a leveled embedding. Some papers have considered *proper* levelings, in which each edge's endpoints are on consecutive levels; we typically do not require our leveling to be proper.

Our results can easily be extended to handle level-planarity testing, an important case of layered graph drawing [6, 13, 14, 16, 19].

Remark 4.3. Level-planar graphs can be recognized and embedded in linear time using PQ-trees [15, 16, 19]; this work is based on earlier work for the special case of hierarchies [6]. There had been an earlier attempt at extending this to general graphs [13, 14], but there were gaps in the algorithm, as pointed out in [17]. Alternative routes have included identifying Kuratowski-style obstruction sets for level-planarity [12], characterizations via vertex-exchange graphs [10, 11], and reductions to 2-satisfiability [29]. It appears that all of these approaches have subtle problems: Currently known obstruction sets for the general case are not complete and are known to be infinite (for standard notions of obstruction containment); only special cases, like trees, are understood [8]. The testing [11] and layout [10] algorithms based on vertex-exchange graphs rest on a characterization of level-planarity that is not fully established at this point, and the case when the vertex-exchange graph is disconnected remains open (P. Healy, January 2011, personal communication); this is unfortunate, since both algorithms are relatively fast, $O(|V|^2)$ for both testing and layout, and very simple (the testing algorithm is somewhat similar to ours, even if the characterization it is based on is different). Finally, there also seems a gap in the suggested reduction to 2-satisfiability (which, if correct, would also result in a quadratic-time testing algorithm).

Thus, although the algorithm we are about to describe may not be the first simple, quadratic-time algorithm for level-planarity testing, it appears to be the first with a complete correctness proof.

Level-planar graphs do not directly generalize x-monotone graphs since an x-monotone graph can have vertices at noninteger levels. However, if G has an x-monotone embedding, then (G,ℓ) is level-planar with $\ell(v) = |\{u : x(u) \leq x(v)\}|$.

Our interest in this section is the reverse direction; how can we reduce testing level-planarity to testing x-monotonicity? The answer is a simple construction: Take a leveled drawing of (G,ℓ). Perturb all vertices slightly, so no two vertices are at the same level. If there is a vertex whose left or right rotation is empty, insert a new edge and vertex on its empty side so that the edges extends slightly beyond all the perturbed vertices from the same level. If there is a vertex with both left and right rotation empty, remove it.

Suppose that the resulting graph G' has an x-monotone embedding with the same vertex locations. By the construction of G', every vertex v that used to have level $\ell(v) = x^*$ is now incident to an edge that passes over the line $x = x^*$. Since all these curves may not intersect each other, we can perturb the drawing slightly (while keeping it x-monotone) to move every vertex of G back to its original level. Also, if (G,ℓ) is level-planar, then G' is obviously x-monotone, so we can use the algorithm from Theorem 4.1 on G' to test level-planarity of (G,ℓ). Since we only added at most $|V(G)|$ vertices and edges to G, the resulting algorithm still runs in quadratic time—with a small constant factor.

Corollary 4.4. *Given a leveled graph (G,ℓ), we can test in time $O(|V|^2)$ whether G is level-planar.*

Note that this result does not require the leveling of G to be proper and thus improves on the algorithm by Healy and Kuusik [11] (assuming it is correct), which requires the leveling to be proper. Turning an improper leveling into a proper leveling (by subdividing edges) can increase the number of vertices by a quadratic factor.

There is one final conclusion we want to draw from the reduction of level planarity to x-monotonicity: When defining a level planar drawing, we required edges to be x-monotone (in the literature, one also finds the equivalent requirement that edges are straight-line segments between levels). As with Corollary 2.7, it is now easy to see that the x-monotonicity requirement is stronger than necessary.

Corollary 4.5. *If (G,ℓ) can be embedded so that $x(v) = \ell(v)$ for every $v \in V$ and every edge is x-bounded, then G is level-planar.*

Proof. Fix an embedding of (G,ℓ) so that $x(v) = \ell(v)$ for every $v \in V$ and every edge is x-bounded. Consider the leveled graph G' constructed before Corollary 4.4. Then G' has a leveled embedding in which every edge is x-bounded. By Corollary 2.7, G' has an x-monotone embedding in which each vertex keeps its x-coordinate (and the rotation system remains unchanged). As above, from this embedding we can obtain a level-planar embedding of G. □

5 Monotone Crossing Numbers

Our Hanani–Tutte results can be recast as results about monotone crossing numbers of leveled graphs. For a leveled graph (G,ℓ), let mon-cr(G,ℓ) be the smallest number of crossings in any leveled drawing of (G,ℓ). Similarly, we can define mon-ocr(G,ℓ) as the smallest number of pairs of edges that cross oddly in any leveled drawing of (G,ℓ). Finally, mon-iocr(G,ℓ) is the smallest number of pairs of nonadjacent edges that cross oddly in any leveled drawing of (G,ℓ). We suppress ℓ and simply write mon-cr(G), mon-ocr(G), and mon-iocr(G). With this notation we can restate the original result by Pach and Tóth, our Theorem 2.11, as saying that mon-ocr$(G) = 0$ implies mon-cr$(G) = 0$. Similarly, our Theorem 3.1 can be restated as mon-iocr$(G) = 0$ implies mon-cr$(G) = 0$.

From this point of view, we can now ask questions that parallel analogous problems for the regular (nonmonotone) crossing number variants cr, ocr, and iocr. For example, we know that $\text{ocr}(G) = \text{cr}(G)$ for $\text{ocr}(G) \leq 3$ [27] and $\text{iocr}(G) = \text{cr}(G)$ for $\text{iocr}(G) \leq 2$ [28]. Pach and Tóth showed that $\text{cr}(G) \leq \binom{2\,\text{ocr}(G)}{2}$ [23,27]. The core step in this result is a "removing even crossings" lemma, in this particular case: If G is drawn in the plane and E_0 is the set of its even edges, then G can be redrawn so that all edges in E_0 are free of crossings. It immediately implies $\text{cr}(G) \leq \binom{2\,\text{ocr}(G)}{2}$, since only noneven edges can be involved in crossings (and every pair of noneven edges needs to cross at most once). A similar result for monotone drawings fails dramatically.

Theorem 5.1. *For every n, there is a graph G so that* mon-cr$(G) \geq n$ *and* mon-ocr$(G) = 1$.

In other words, even if there are only two edges crossing each other oddly and all other edges are even, then any x-monotone drawing of G with the given leveling may require an arbitrary number of crossings. Thus, we cannot hope to establish a "removing even crossings" lemma in the context of x-monotone drawings since it would imply a bound on mon-cr(G) in terms of mon-ocr(G).

Proof. Our example uses 8 vertices, allowing multiple edges, which we bundle into a single weighted edge. Consider the graph on 8 vertices with edges 36 and 57 of weight 1 and edges 12, 13, 25, 26, 37, 46, 47, 68, and 78 of weight $n > 1$. Weighted edges can be replaced by paths of length 2, turning the example into a simple graph.

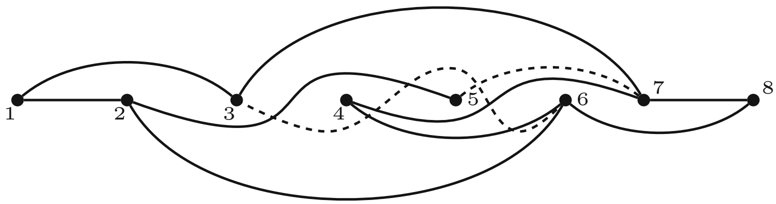


Fig. 8 The drawing showing mon-ocr$(G) \leq 1$. The *solid edges* have weight n, and the *dashed edges* have weight 1

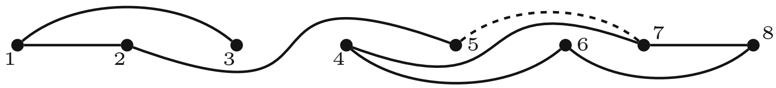


Fig. 9 The unique way of drawing edges 25, 57, 68, and 27, assuming 46 passes below 5 and mon-cr$(D) < n$

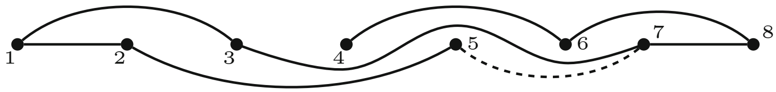


Fig. 10 The unique way of drawing edges 25, 57, 68, and 37, assuming 46 passes above 5 and mon-cr$(D) < n$

We next argue that mon-cr$(G) \geq n$. Suppose there is a drawing D with mon-cr$(D) < n$. Then the only pair of edges that may intersect is 36 and 57. Without loss of generality, we can assume that 12, 13 and 78 are drawn exactly as they are in Fig. 8. We distinguish two cases depending on whether 46 passes below 5 (as in Fig. 8) or above 5. Let's first consider the case that 46 passes below 5. Adding edges 25, 57, we see that they are forced to be drawn as in Fig. 8. At this point, edge 68 has to pass below 7 and then 47 is forced. That is, if we assume that 46 passes below 5, then the edges we added have to be drawn as shown in Fig. 9. By inspection, it is clear that adding edge 36 to this drawing will cause at least n crossings, either with edge 25 or with edge 47.

On the other hand, if 46 passes above 5, then edge 25 is forced to pass below 3 and 4 and edge 57 is forced below 6. This forces 68 above 7, which in turn forces 37 below 4 and 6 and above 5. However, now it is impossible to add edge 26 without having it cross either 13 or 37; see Fig. 10. □

6 Future Directions

The following questions were first included in the conference version of this chapter. Since that time, there has already been some progress, which we include as annotations.

Planarity testing Can the running time of our level-planarity testing algorithms be improved? There are several obstacles to this, most fundamentally the problem that to describe the effect of (e, v)-moves, we need a system with a quadratic number of variables. It is not obvious (to us at least) how to reduce the size of this system. Other problems in the algorithm, such as the linear overhead in the "conquering" steps of the algorithm, may be overcome with better data structures.

Monotone crossing numbers The *monotone crossing number* of a leveled graph G is the smallest number of crossings in any x-monotone drawing of the leveled graph. This problem is known to be **NP**-hard (even for two levels [9]), and the monotone crossing number can be arbitrarily large, even for a planar graph.[8]

We get a more interesting question if we define the *monotone crossing number for unleveled graphs* as the smallest crossing number of any x-monotone drawing for any leveling of the graph. Is this monotone crossing number bounded in the crossing number? For comparison, $\mathrm{rcr}_2(G)$ is at most $\binom{\mathrm{cr}(G)}{2}$, where $\mathrm{rcr}_2(G)$ allows straight-line edges with one bend [1].

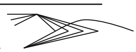
Pach and Tóth [26] recently announced that the (unleveled) monotone crossing number of a graph G can be bounded by $2\,\mathrm{cr}(G)^2$ and that there are graphs for which the monotone crossing number is at least $7/6\,\mathrm{cr}(G) - 6$. We should also mention that Valtr [32] showed that the monotone crossing number is bounded by $4k^{4/3}$, where k is the monotone pair crossing number (again for unleveled graphs).

Bi-monotonicity Let's define y-monotonicity like x-monotonicity after a 90-degree rotation; not very exciting by itself, but what happens if we want embeddings that are *bi-monotone*, that is, both x- and y-monotone?

- If a graph has both an x-monotone embedding and a y-monotone embedding, does it always have a bi-monotone embedding?
- If there is a drawing of a graph that is bi-monotone, is there a straight-line drawing with the same x and y ordering?
- What about bilevel-planarity?

As far as we know, bi-monotonicity and bilevel-planarity are new concepts; however, they are quite natural: If we specify the relative locations of objects on a map, we specify them in terms of "west/east of" and "north/south of," which is exactly what bi-monotonicity models. Imagine specifying the stations for a subway map: Actual distances do not matter; what matters is relative location in terms of x and y.

As it turns out, it is possible that a leveled graph has both an x-monotone and a y-monotone embedding without having a bi-monotone embedding; see the leveled graph in Fig. 11. By Theorem 2.11, the graph is x-monotone and

[8] The leveled graph  is such an example.

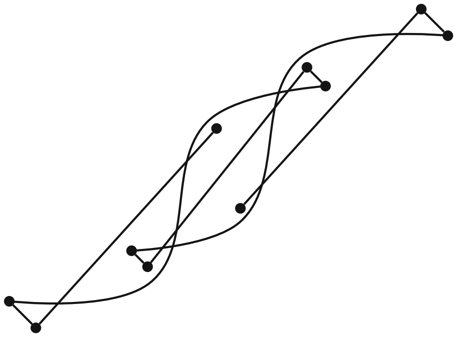


Fig. 11 A leveled path that is not bi-monotone

(applying the theorem at an angle of 90 degrees) *y*-monotone. However, it can be shown that the graph is not bi-monotone.

This means that an algebraic bi-monotonicity criterion has to be more sophisticated than just requiring a bi-monotone even drawing. It also opens up the question of what is the complexity of recognizing bi-monotone or bilevel planar graphs?

Acknowledgements We thank the anonymous referee for careful proofreading.

Radoslav Fulek gratefully acknowledges support from the Swiss National Science Foundation Grant No. 200021-125287/1.

Michael J. Pelsmajer gratefully acknowledges the support from NSA Grant H98230-08-1-0043 and the Swiss National Science Foundation Grant No. 200021-125287/1.

References

1. D. Bienstock, N. Dean, Bounds for rectilinear crossing numbers. J. Graph Theor. **17**(3), 333–348 (1993)
2. G. Cairns, Y. Nikolayevsky, Bounds for generalized thrackles. Discr. Comput. Geom. **23**(2), 191–206 (2000)
3. C. Chojnacki (Haim Hanani), Über wesentlich unplättbare Kurven im drei-dimensionalen Raume. Fund. Math. **23**, 135–142 (1934)
4. H. de Fraysseix, Trémaux trees and planarity, in *The International Conference on Topological and Geometric Graph Theory*, Electronics Notes in Discrete Mathematics, vol. 31 (Elsevier Science B. V., Amsterdam, 2008), pp. 169–180
5. H. de Fraysseix, P. Rosenstiehl, A characterization of planar graphs by Trémaux orders. Combinatorica **5**(2), 127–135 (1985)
6. G. Di Battista, E. Nardelli, Hierarchies and planarity theory. IEEE Trans. Syst. Man Cybernet. **18**(6), 1035–1046 (1988/1989)
7. P. Eades, Q. Feng, X. Lin, H. Nagamochi, Straight-line drawing algorithms for hierarchical graphs and clustered graphs. Algorithmica **44**(1), 1–32 (2006)
8. A. Estrella-Balderrama, J. Joseph Fowler, S.G. Kobourov, On the characterization of level planar trees by minimal patterns, in *Graph Drawing*, ed. by D. Eppstein, E.R. Gansner. Lecture Notes in Computer Science, vol. 5849 (Springer-Verlag, Berlin, 2009), pp. 69–80
9. M.R. Garey, D.S. Johnson, Crossing number is NP-complete. SIAM J. Algebraic Discr. Methods **4**(3), 312–316 (1983)

10. M. Harrigan, P. Healy, Practical level planarity testing and layout with embedding constraints, in *Graph Drawing*. Lecture Notes in Computer Science, vol. 4875 (Springer-Verlag, 2008), pp. 62–68
11. P. Healy, A. Kuusik, Algorithms for multi-level graph planarity testing and layout. Theor. Comput. Sci. **320**(2–3), 331–344 (2004)
12. P. Healy, A. Kuusik, S. Leipert, A characterization of level planar graphs. Discr. Math. **280** (1–3), 51–63 (2004)
13. L.S. Heath, S.V. Pemmaraju, Recognizing leveled-planar dags in linear time, in *Graph Drawing, GD'95*, ed. by F.J. Brandenburg. Lecture Notes in Computer Science, vol. 1027 (Springer-Verlag, Berlin, 1996), pp. 300–311
14. L.S. Heath, S.V. Pemmaraju, Stack and queue layouts of directed acyclic graphs. II. SIAM J. Comput. **28**(5), 1588–1626 (1999) (electronic)
15. M. Jünger, S. Leipert, Level planar embedding in linear time. J. Graph Algorithm. Appl. **6**(1), 67–113 (2002) (electronic)
16. M. Jünger, S. Leipert, P. Mutzel, Level planarity testing in linear time, in *Graph Drawing*, Montréal, QC, 1998. Lecture Notes in Computer Science, vol. 1547 (Springer-Verlag, Berlin, 1998), pp. 224–237
17. M. Jünger, S. Leipert, P. Mutzel, Pitfalls of using *PQ*-trees in automatic graph drawing, in *Proceedings of the 5th International Symposium on Graph Drawing, GD'97*, Rome, September 18–20, 1997, ed. by G. DiBattista. LNCS, vol. 1353 (Springer-Verlag, Berlin, 1998), pp. 193–204
18. D.J. Kleitman, A note on the parity of the number of crossings of a graph. J. Comb. Theor. Ser. B **21**(1), 88–89 (1976)
19. S. Leipert, Level planarity testing and embedding in linear time. Ph.D. thesis, Universität zu Köln, Köln, 1998
20. X. Lin, P. Eades, Towards area requirements for drawing hierarchically planar graphs. Theor. Comput. Sci. **292**(3), 679–695 (2003)
21. J. Matoušek, *Using the Borsuk–Ulam theorem. Universitext* (Springer, Berlin, 2003)
22. J. Matousek, M. Tancer, U. Wagner, Hardness of embedding simplicial complexes in $\mathbb{R}^d$, in *Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2009*, New York, January 4–6, 2009, ed. by C. Mathieu (SIAM, Philadelphia, PA, 2009), pp. 855–864
23. J. Pach, G. Tóth, Which crossing number is it anyway? J. Comb. Theor. Ser. B **80**(2), 225–246 (2000)
24. J. Pach, G. Tóth, Monotone drawings of planar graphs. J. Graph Theor. **46**(1), 39–47 (2004)
25. J. Pach, G. Tóth, Monotone drawings of planar graphs. ArXiv:1101.0967 e-prints (January 2011)
26. J. Pach, G. Tóth, Monotone crossing number, in *Proceedings of the 19th International Symposium on Graph Drawing, GD'11*, Eindhoven, September 21–23, 2011, ed. by Marc J. Van Kreveld and Bettina Speckmann. LNCS, **7034** (Springer-Verlag, Berlin, 2012), pp. 278–289
27. M.J. Pelsmajer, M. Schaefer, D. Štefankovič, Removing even crossings. J. Comb. Theor. Ser. B **97**(4), 489–500 (2007)
28. M.J. Pelsmajer, M. Schaefer, D. Štefankovič, Removing independently even crossings. SIAM J. Discr. Math. **24**(2), 379–393 (2010)
29. B. Randerath, E. Speckenmeyer, E. Boros, P.L. Hammer, A. Kogan, K. Makino, B. Simeone, O. Cepek, A satisfiability formulation of problems on level graphs. Electron. Notes Discr. Math. **9**, 269–277 (2001)
30. M. Schaefer, Hanani–Tutte and related results, in *Geometry-Instuitive, Discrete and Convex-A Tribute to László Fejes Tóth* (I.Bárány, K.J. Böröczky, G. Fejes Tóth, J. Pach, Eds.), Bolyai society Mathematical Studies, (Springer, Berlin) to apper
31. W.T. Tutte, Toward a theory of crossing numbers. J. Comb. Theor. **8**, 45–53 (1970)
32. P. Valtr, On the pair-crossing number, in *Combinatorial and Computational Geometry*. Math. Sci. Res. Inst. Publ., vol. 52 (Cambridge University Press, Cambridge, 2005), pp. 569–575

33. W.J. Wu, On the planar imbedding of linear graphs. I. J. Syst. Sci. Math. Sci. **5**(4), 290–302 (1985)
34. W.J. Wu, On the planar imbedding of linear graphs (continued). J. Syst. Sci. Math. Sci. **6**(1), 23–35 (1986)

On Disjoint Crossing Families in Geometric Graphs

Radoslav Fulek and Andrew Suk

Abstract A *geometric graph* is a graph drawn in the plane with vertices represented by points and edges as straight-line segments. A geometric graph contains a (k,l)-*crossing family* if there is a pair of edge subsets E_1, E_2 such that $|E_1| = k$ and $|E_2| = l$, the edges in E_1 are pairwise crossing, the edges in E_2 are pairwise crossing, and every edge in E_1 is disjoint to every edge in E_2. We conjecture that for any fixed k, l, every n-vertex geometric graph with no (k,l)-crossing family has at most $c_{k,l}n$ edges, where $c_{k,l}$ is a constant that depends only on k and l. In this note, we show that every n-vertex geometric graph with no (k,k)-crossing family has at most $c_k n \log n$ edges, where c_k is a constant that depends only on k, by proving a more general result that relates an extremal function of a geometric graph F with an extremal function of two completely disjoint copies of F. We also settle the conjecture for geometric graphs with no $(2,1)$-crossing family. As a direct application, this implies that for any circle graph F on three vertices, every n-vertex geometric graph that does not contain a matching whose intersection graph is F has at most $O(n)$ edges.

1 Introduction

A *topological graph* is a graph drawn in the plane with points as vertices and edges as nonself-intersecting arcs connecting its vertices. Two edges of a topological graph *cross* if their interiors share a point and are *disjoint* if they do not have a point in common (including their endpoints). The edges are allowed to intersect,

R. Fulek (✉)
Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
e-mail: radoslav.fulek@epfl.ch

A. Suk
Department of Mathematics, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA
e-mail: asuk@math.mit.edu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_15,

but they may not pass through vertices. Furthermore, the edges are not allowed to have tangencies; i.e., if two edges share an interior point, then they must properly cross at that point. We only consider graphs without parallel edges or self-loops. A topological graph is *simple* if every pair of its edges intersects at most once. If the edges are drawn as straight-line segments, then the graph is *geometric*.

It follows from Euler's polyhedral formula that every simple topological graph on $n \geq 3$ vertices and no crossing edges has at most $3n-6$ edges. It is also known that every simple topological graph on n vertices with no pair of disjoint edges has at most $O(n)$ edges [7, 10]. Finding the maximum number of edges in a topological (and geometric) graph with a forbidden substructure has been a classic problem in extremal topological graph theory (see [1, 2, 6, 13, 14, 17–20]). Many of these problems ask for the maximum number of edges in a topological (or geometric) graph whose edge set does not contain a matching that defines a particular intersection graph. Recall that the *intersection graph* of a set of objects $\mathcal{C}$ in the plane is a graph with vertex set $\mathcal{C}$, and two vertices are adjacent if their corresponding objects intersect. Much research has been devoted to understanding the clique and independence number of intersection graphs due to their applications in VLSI design [8], map labeling [3], and elsewhere.

Recently, Ackerman et al. [4] defined a *natural* (k,l)*-grid* to be a set of k pairwise disjoint edges that all cross another set of l pairwise disjoint edges. They conjectured

Conjecture 1.1. Given fixed constants $k,l \geq 1$, there exists another constant $c_{k,l}$ such that any geometric graph on n vertices with no natural (k,l)-grid has at most $c_{k,l}n$ edges.

They were able to show

Theorem 1.2 [4]. *For fixed k, an n-vertex geometric graph with no natural* (k,k)*-grid has at most* $O(n\log^2 n)$ *edges.*

Theorem 1.3 [4]. *An n-vertex geometric graph with no natural* $(2,1)$*-grid has at most* $O(n)$ *edges.*

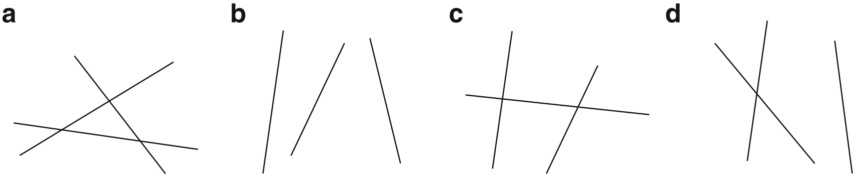
Theorem 1.4 [4]. *An n-vertex simple topological graph with no natural* (k,k)*-grid has at most* $O(n\log^{4k-6} n)$ *edges.*

As a *dual* version of the natural (k,l)-grid, we define a (k,l)*-crossing family* to be a pair of edge subsets E_1, E_2 such that

1. $|E_1| = k$ and $|E_2| = l$.
2. The edges in E_1 are pairwise crossing.
3. The edges in E_2 are pairwise crossing.
4. Every edge in E_1 is disjoint to every edge in E_2.

We conjecture the following.

Conjecture 1.5. Given fixed constants $k,l \geq 1$ there exists another constant $c_{k,l}$ such that any geometric graph on n vertices with no (k,l)-crossing family has at most $c_{k,l}n$ edges.



a 3 pairwise crossing. b 3 pairwise disjoint. c (2,1)-grid. d (2,1)-crossing family.

Fig. 1 Triples of segments corresponding to all graphs on three vertices

It is not even known if all n-vertex geometric graphs with no k pairwise crossing edges has $O(n)$ edges. The best-known bound is due to Valtr [21], who showed that this is at most $O(n\log n)$ for every fixed k. We extend this result to (k,k)-crossing families by proving the following theorem.

Theorem 6. *An n-vertex geometric graph with no (k,k)-crossing family has at most $c_k n\log n$ edges, where c_k is a constant that depends only on k.*

Let F denote a geometric graph. We say that a geometric graph G contains F as a geometric subgraph if G contains a subgraph F' isomorphic to F such that two edges in F' cross if and only if the two corresponding edges cross in F.

We define $ex(F,n)$ to be the extremal function of F, i.e., the maximum number of edges a geometric graph on n vertices can have without containing F as a geometric subgraph. Similarly, we define $ex_L(F,n)$ to be the extremal function of F if we restrict ourselves to the geometric graphs all of whose edges can be hit by one line.

Let F_2 denote a geometric graph that consists of two completely disjoint copies of a geometric graph F. We prove Theorem 6 by a straightforward application of the following result.

Theorem 7. $ex(F_2,n) = O((ex_L(F,2n)+n)\log n + ex(F,n))$.

Furthermore, we settle Conjecture 1.5 in the first nontrivial case.

Theorem 8. *An n-vertex geometric graph with no $(2,1)$-crossing family has at most $O(n)$ edges.*

Note that Conjecture 1.5 is not true for topological graphs since Pach and Tóth [15] showed that the complete graph can be drawn such that every pair of edges intersect once or twice.

By combining Theorem 8 with results from [2, 4, 19], we have the following.

Corollary 1.9. *For any graph F on 3 vertices, every n-vertex geometric graph that does not contain a matching whose intersection graph is F contains at most $O(n)$ edges.*

See Fig. 1. Recall that F is a *circle graph* if F can be represented as the intersection graph of chords on a circle. Corollary 1.9 cannot be generalized to all graphs F, because if the vertices of a geometric graph G are in convex position, the intersection graph of any subset of edges of G is a circle graph. Hence, we conjecture the following.

Conjecture 1.10. For any circle graph F on k vertices, there exists a constant c_k such that every n-vertex geometric graph that does not contain a matching whose intersection graph is F contains at most $c_k n$ edges.

As pointed out by Klazar and Marcus [9], it is not hard to modify the proof of the Marcus–Tardos theorem [12] to show that Conjecture 1.10 is true when the vertices are in convex position.

For simple topological graphs, we have the following.

Theorem 11. *An n-vertex simple topological graph with no $(k,1)$-crossing family has at most $n(\log n)^{O(\log k)}$ edges.*

The chapter is organized as follows. Section 2 is devoted to the proof of Theorem 7. In Sect. 3 we establish Theorem 8. Finally, the result of Theorem 11 about topological graphs is proved in Sect. 4.

2 Relating Extremal Functions

First, we prove a variant of Theorem 7 when all of the edges in our geometric graph can be hit by a line. As in the Introduction, let F_2 denote a geometric graph that consists of two completely disjoint copies of a geometric graph F. We will now show that the extremal function $ex_L(F_2,n)$ is not far from $ex_L(F,n)$.

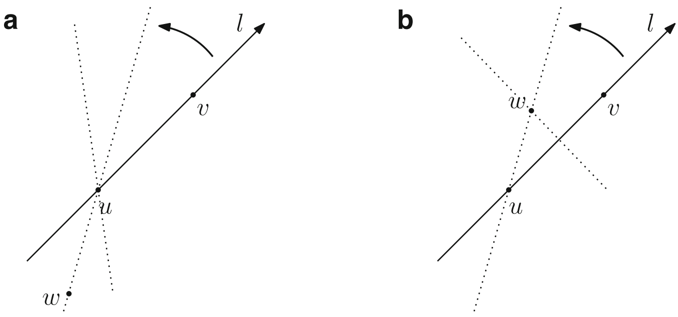
Theorem 1. $ex_L(F_2,n) \leq O((n+ex_L(F,2n))\log n)$.

Proof. Let G denote a geometric graph on n vertices that does not contain F_2 as a geometric subgraph, and all the edges of G can be hit by a line. By a standard perturbation argument, we can assume that the vertices of G are in general position. As in [5], a *halving edge uv* is a pair of the vertices in G such that the number of vertices on each side of the line through u and v is the same. By a *halving line*, we understand a line that bisects the set of vertices of G.

Lemma 2.2. *There exists a directed halving line $\vec{l}$ such that the number of edges in G that lies completely to the left or right of $\vec{l}$ is at most $2ex_L(F,n/2)+7n$.*

Proof. If n is odd, we can discard one vertex of G, thereby losing at most n edges. Therefore, we can assume n is even, and it suffices to show that there exists a directed halving line $\vec{l}$ such that the number of edges in G that lies completely to the left or right of $\vec{l}$ is at most $2ex_L(F,n/2)+4n$.

Let uv be a halving edge, and let $\vec{l}$ denote the directed line containing vertices u and v with direction from u to v. Let $e(\vec{l},L)$ and $e(\vec{l},R)$ denote the number of edges on the left and right side of $\vec{l}$, respectively. Without loss of generality, we can assume that $e(\vec{l},L) \leq e(\vec{l},R)$. We will rotate $\vec{l}$ such that it remains almost a halving line at



w lies to the right of $\vec{l}$.

w lies to the left of $\vec{l}$.

Fig. 2 Halving the vertices of G

the end of each step, until it reaches a position where the number of edges on both sides of $\vec{l}$ is roughly the same.

We start by rotating $\vec{l}$ counterclockwise around u until it meets the next vertex w of G. If initially w lies to the right of $\vec{l}$, then in the next step we will rotate $\vec{l}$ around u (again). See Fig. 2a. Otherwise, if w was on the left side of $\vec{l}$, then in the next step we will rotate $\vec{l}$ around vertex w. See Fig. 2b. Clearly, during the rotation there are always at least $n/2-2$ and at most $n/2$ vertices on each side of $\vec{l}$.

After several rotations, $\vec{l}$ will eventually contain points u and v again, with direction from v to u. Indeed, after rotating $\vec{l}$ by 180°, its slope will be the same as the slope of its initial position. Hence, if $\vec{l}$ in this position does not pass through u and v, either its right or left side contains at most $n/2-3$ points (or at least $n/2+1$). At this point we have $e(\vec{l},L) \geq e(\vec{l},R)$. Since the number of edges on the right side (and left side) changes by at most n after each step in the rotation, at some point in the rotation we must have

$$|e(\vec{l},L) - e(\vec{l},R)| \leq 2n.$$

Since G does not contain F_2 as a geometric subgraph, this implies that

$$e(\vec{l},L) \leq ex_L(F,n/2) + 2n$$

and

$$e(\vec{l},R) \leq ex_L(F,n/2) + 2n.$$

Therefore, for any n, there exists a directed halving line $\vec{l}$ such that the number of edges in G that lies completely to the left or right of $\vec{l}$ is at most $2ex_L(F,n/2)+7n$. □

By Lemma 2.2 we obtain a line l, which partitions the vertices of G into two equal (or almost equal if n is odd) sets V_1 and V_2. Let E' denote the set of edges

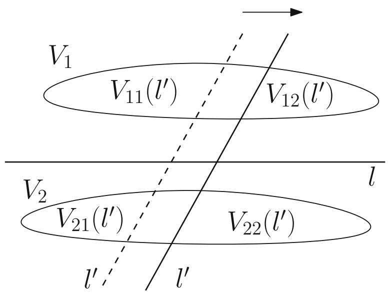


Fig. 3 The final partition of the vertex set of G

between V_1 and V_2. By the ham-sandwich cut theorem [11], there exists a line l' that simultaneously bisects V_1 and V_2. Let $V_{11}(l')$ and $V_{12}(l')$ denote the resulting parts of V_1, and let $V_{21}(l')$ and $V_{22}(l')$ denote the resulting parts of V_2.

Observe that we can translate l' along l into a position where the number of edges in E' that lie completely to the left and completely to the right of l' is roughly the same. In particular, we can translate l' along l such that the number of edges in E' that lies completely to its left or right side is at most $ex_L(F,n)+ex_L(F,n/2+1)+n$ (see Fig. 3). Indeed, assume that the number of edges in E' between, say, $V_{12}(l')$ and $V_{22}(l')$ is more than $ex_L(F,n/2+1)$. As we translate l' to the right, the number of edges that lie completely to the right of l' changes by at most n as l' passes through a single vertex in G. Therefore, we can translate l' into the leftmost position, where the number of edges in E' between $V_{12}(l')$ and $V_{22}(l')$ drops below $ex_L(F,n/2+1)+n+1$. Moreover, at this position the number of edges in E' between $V_{11}(l')$ and $V_{21}(l')$ still cannot be more than $ex_L(F,n)$ since G does not contain F_2 as a geometric subgraph.

Thus, all but at most $3ex_L(F,n/2+1)+ex_L(F,n)+8n$ edges of G are the edges between $V_{11}(l')$ and $V_{22}(l')$, and between $V_{12}(l')$ and $V_{21}(l')$. Since $|V_{11}(l')|,|V_{21}(l')| \geq \lfloor (n-1)/4 \rfloor$, there exists k, $-1/4 \leq k \leq 1/4$, such that $|V_{11}(l')|+|V_{22}(l')| = n(1/2+k) \pm O(1)$, and $|V_{12}(l')|+|V_{21}(l')| = n(1/2-k) \pm O(1)$. Finally, we are in the position to state the recurrence, whose closed form gives the statement of the theorem (n_0 and c are universal constants):

$$\begin{aligned} ex_L(F_2,n) &\leq c,\ n \leq n_0 \\ ex_L(F_2,n) &\leq ex_L(F_2,n(1/2+k)) + ex_L(F_2,n(1/2-k)) + 3ex_L(F,n/2+1) + \\ &\qquad ex_L(F,n) + 8n + O(1),\ n > n_0. \end{aligned} \tag{1}$$

By plugging into (1) and using the superadditivity of ex_L, we verify that for sufficiently large n, we have $ex_L(F_2,n) \leq \log_{\frac{4}{3}} n(4ex_L(F,2n)+9n)$:

$$ex_L(F_2,n) \leq \log_{\frac{4}{3}}\left(n\left(\frac{1}{2}+k\right)\right)\left(9n\left(\frac{1}{2}+k\right)+4ex_L\left(F,2n\left(\frac{1}{2}+k\right)\right)\right)$$

$$
\begin{aligned}
&+\log_{\frac{4}{3}}\left(n\left(\frac{1}{2}-k\right)\right)\left(9n\left(\frac{1}{2}-k\right)+4ex_L\left(F,2n\left(\frac{1}{2}-k\right)\right)\right)\\
&+4ex_L(F,n)+8n+O(1)\\
\leq\ &\log_{\frac{4}{3}}\left(\frac{3}{4}n\right)\left(9n\left(\frac{1}{2}+k\right)+4ex_L\left(F,2n\left(\frac{1}{2}+k\right)\right)\right)\\
&+\log_{\frac{4}{3}} n\left(9n\left(\frac{1}{2}-k\right)+4ex_L\left(F,2n\left(\frac{1}{2}-k\right)\right)\right)\\
&+4ex_L(F,n)+8n+O(1)\\
\leq\ &\log_{\frac{4}{3}} n(4ex_L(F,2n)+9n). \qquad \square
\end{aligned}
$$

Finally, we show how Theorem 1 implies Theorem 7.

Proof of Theorem 7. Let $G=(V,E)$ denote the geometric graph not containing F_2 as a subgraph. Similarly, as in the proof of Lemma 2.2, we can find a halving line l that hits all but $2ex(F,n/2)+9n$ edges of G. Now, the claim follows by using Theorem 1. $\square$

Theorem 6 follows easily by using Theorem 7 with a result from [21], which states that every n-vertex geometric graph whose edges can be all hit by a line and does not contain k pairwise crossing edges has at most $O(n)$ edges and at most $O(n\log n)$ edges if we do not require a single line to hit all the edges.

3 Geometric Graphs with No (2, 1)-Crossing Family

In this section, we will prove Theorem 8. Our main tool is the following theorem by Tóth and Valtr.

Theorem 1 ([20]). *Let $G=(V,E)$ be an n-vertex geometric graph. If G does not contain a matching consisting of 5 pairwise disjoint edges, then $|E(G)|\leq 64n+64$.*

Theorem 8 immediately follows from the following theorem.

Theorem 2. *Every n-vertex geometric graph with no $(2,1)$-crossing family has at most $64n+64$ edges.*

Proof. For the sake of contradiction, let $G=(V,E)$ be a vertex-minimal counterexample; i.e., G is a geometric graph on n vertices that has more than $64n+64$ edges and G does not contain a (2,1)-crossing family. Hence, every vertex in G has degree at least 65. Let M denote the maximum matching in G consisting of pairwise disjoint edges, and let V_M denote the vertices in M. Since $|E(G)|>64n+64$, Theorem 1 implies that $|M|\geq 5$. We say that two edges *intersect* if they cross or share an endpoint. The following simple observation is crucial in

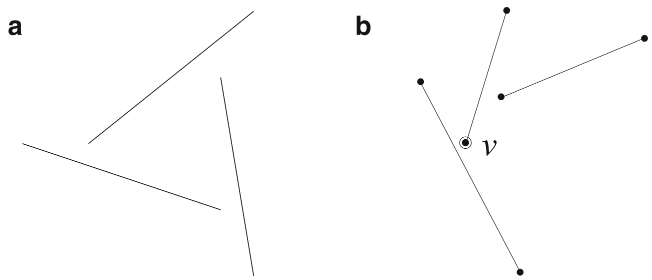


Fig. 4 (**a**) Special case in Lemma 3.3; (**b**) matching M of size 3 with one good vertex v

the subsequent analysis.

(*) An edge $e \in E$ that crosses an edge of M must intersect every edge in M.

Indeed, otherwise we would obtain a (2,1)-crossing family. We call an endpoint v of an edge in M *good* if every ray starting at v misses at least one edge in M. See Fig. 4b.

Lemma 3.3. *For $|M| \geq 4$, at least $|M|-3$ of the endpoints in M are good.*

Proof. We proceed by induction on $|M|$. The only matching consisting of three pairwise disjoint edges with no good vertices is the one in Fig. 4a. Thus, adding a disjoint edge to this matching creates at least one good vertex. For the inductive step when $|M| > 4$, we choose an arbitrary 4-tuple of edges in the matching. By the above discussion, the 4-tuple has at least one good endpoint. By removing the edge incident to this good vertex, the statement follows by the induction hypothesis. □

Thus, by (*), a good endpoint cannot be incident to an edge that crosses any of the edges in M. Let

1. $V_g \subseteq V_M$ denote the set of good endpoints in M.
2. $V_1 \subset V \setminus V_M$ be the subset of the vertices such that for $v \in V_1$, every edge incident to v does not cross any of the edges in M.
3. $V_2 = V \setminus (V_M \cup V_1)$. Hence, for $v \in V_2$, there exists an edge incident to v that intersects every edge in M.

See Fig. 5a. By Lemma 3.3, $|V_g| \geq |M| - 3 = |V_M|/2 - 3$. By the maximality of M, there are no edges between V_1 and V_2 and V_1 is an independent set. Now notice the following observation.

Observation 3.4. *There exists a good vertex in V_g that has at least three neighbors in V_2.*

Proof. For the sake of contradiction, suppose that each vertex in V_g has at most two neighbors in V_2. Then let $G' = (V', E')$ denote a subgraph of G such that $V' = V_M \cup V_1$ and E' consists of the edges that do not cross any of the edges of M and whose

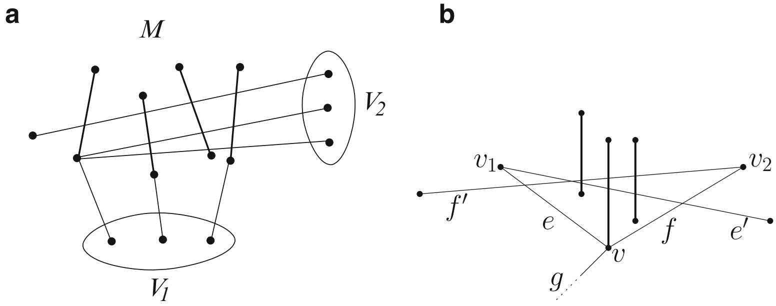


Fig. 5 (**a**) M, V_1, and V_2; (**b**) situation around the vertex v

endpoints are in $V_1 \cup V_M$. Since $|M| \geq 5$, G' must be a planar graph since otherwise we would have a $(2,1)$-crossing family. Therefore, $|E'| \leq 3(|V_1| + |V_M|)$.

On the other hand, by the minimality of G, each vertex in V_1 has degree at least 65 in G', and each vertex in V_g has degree at least 63 in G'. Therefore, by applying Lemma 3.3, we have

$$\frac{1}{2}\left(65|V_1| + 63\left(\frac{|V_M|}{2} - 3\right)\right) \leq |E'| \leq 3|V_1| + 3|V_M|.$$

This implies

$$59|V_1| + 25|V_M| \leq 189,$$

which is a contradiction since $|V_M| \geq 10$ ($|M| \geq 5$). □

Let $v \in V_g$ be a good vertex such that v has at least three neighbors in V_2. Let $e = vv_1$, $f = vv_2$, $g = vv_3$, and m be edges in G such that $v_1, v_2, v_3 \in V_2$, $m \in M$, and v is a good vertex incident to m. Furthermore, we will assume that g, e, m, f appear in clockwise order around v. By (*) there is an edge e' incident to v_1 having a nonempty intersection with every edge in M. Similarly, we can find such an edge f' incident to v_2 (possibly $f' = e'$). The edges e' and f' must have a nonempty intersection with f and e, respectively (see Fig. 5b). Otherwise, we would obtain a (2,1)-crossing family in G consisting of e, f' and an edge from M, or e', f and an edge from M.

However, a third edge g cannot have a nonempty intersection with both e' and f'. Hence, we obtain a (2,1)-crossing family in G consisting of g, e' and an edge from M, or g, f' and an edge from M. Thus, there is no minimal counterexample and that concludes the proof. □

We note that by a more tedious case analysis, one could improve the upper bound in Theorem 2 to $15n$.

4 Simple Topological Graphs with No $(k,1)$-Crossing Family

In this section, we will prove Theorem 11, which will require the following two lemmas. The first one is due to Fox and Pach.

Lemma 4.1 ([6]). *Every n-vertex simple topological graph with no k pairwise crossing edges has at most* $n(\log n)^{c_1 \log k}$ *edges, where* c_1 *is an absolute constant.*

As defined in [16], the *odd-crossing number* odd-cr(G) of a graph G is the minimum possible number of unordered pairs of edges that cross an odd number of times over all drawings of G. The *bisection width* of a graph G, denoted by $b(G)$, is the smallest nonnegative integer such that there is a partition of the vertex set $V = V_1 \dot\cup V_2$ with $\frac{1}{3}\cdot|V| \le V_i \le \frac{2}{3}\cdot|V|$ for $i=1,2$, and $|E(V_1,V_2)| = b(G)$. The second lemma required is due to Pach and Tóth, which relates the odd-crossing number of a graph to its bisection width.

Lemma 4.2 ([15]). *There is an absolute constant* c_2 *such that if G is a graph with n vertices of degrees* $d_1,\ldots,d_n$*, then*

$$b(G) \le c_2 \log n \sqrt{\text{odd-cr}(G) + \sum_{i=1}^{n} d_i^2}.$$

Since all graphs have a bipartite subgraph with at least half of its edges, Theorem 6 immediately follows from the following theorem.

Theorem 3. *Every n vertex simple topological bipartite graph with no* $(k,1)$*-crossing family has at most* $c_3 n \log^{c_4 \log k} n$ *edges, where* c_3, c_4 *are absolute constants.*

Proof. We proceed by induction on n. The base case is trivial. For the inductive step, the proof falls into two cases.

Case 1. Suppose there are at least $|E(G)|^2/((2c_2)^2\log^6 n)$ disjoint pairs of edges in G. Then by defining $D(e)$ to be the set of edges disjoint from edge e, we have

$$\frac{2|E(G)|}{(2c_2)^2\log^6 n} \le \frac{\sum\limits_{e\in E(G)} |D(e)|}{|E(G)|}.$$

Hence, there exists an edge that is disjoint to at least $2|E(G)|/((2c_2)^2\log^6 n)$ other edges. By Lemma 4.1, we have

$$\frac{2|E(G)|}{(2c_2)^2\log^6 n} \le n(\log n)^{c_1 \log k},$$

which implies $|E(G)| \le c_3 n \log^{c_4 \log k} n$ for sufficiently large constants c_3, c_4.

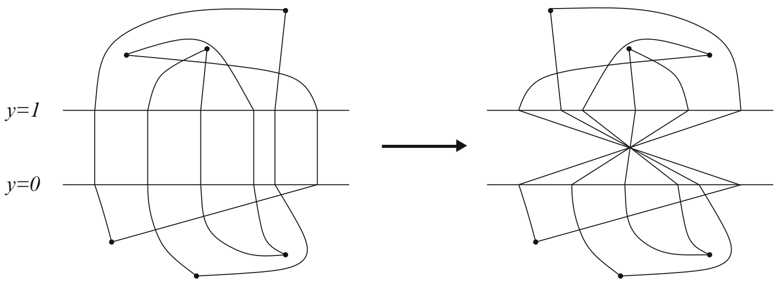


Fig. 6 Redrawing procedure

Case 2. Suppose there are at most $|E(G)|^2/((2c_2)^2 \log^6 n)$ disjoint pairs of edges in G. In what follows, we will apply a redrawing technique that was used by Pach and Tóth [15]. Since G is bipartite, let V_a and V_b be its vertex class. By applying a suitable homeomorphism to the plane, we can redraw G such that

1. The vertices in V_a are above the line $y = 1$, the vertices in V_b are below the line $y = 0$.
2. Edges in the strip $0 \leq y \leq 1$ are vertical segments.
3. We have neither created nor removed any crossings.

Now we reflect the part of G that lies above the $y = 1$ line about the y-axis. Then erase the edges in the strip $0 \leq y \leq 1$ and replace them by straight-line segments that reconnect the corresponding pairs on the line $y = 0$ and $y = 1$. See Fig. 6, and note that our graph is no longer simple.

Notice that if any two edges crossed in the original drawing, then they must cross an even number of times in the new drawing. Indeed, suppose the edges e_1 and e_2 crossed in the original drawing. Since G is simple, they share exactly one point in common. Let k_i denote the number of times edge e_i crosses the strip for $i \in \{1,2\}$, and note that k_i must be odd. After we have redrawn our graph, these $k_1 + k_2$ segments inside the strip will now pairwise cross, creating $\binom{k_1+k_2}{2}$ crossing points. Since edge e_i will now cross itself $\binom{k_i}{2}$ times, this implies that there are now

$$\binom{k_1+k_2}{2} - \binom{k_1}{2} - \binom{k_2}{2} \tag{2}$$

crossing points between edges e_1 and e_2 inside the strip. One can easily check that (2) is odd when k_1 and k_2 are odd. Since e_1 and e_2 had one point in common outside the strip, this implies that e_1 and e_2 cross each other an even number of times. Note that one can easily get rid of self-intersections by making local modifications around these crossings.

Hence, the odd-crossing number in our new drawing is at most the number of noncrossing pair of edges in the original drawing of G. Since there are at most

$\sum_{v \in V(G)} d^2(v) \leq 2|E(G)|n$ pairs of edges that share a vertex in G, this implies

$$\text{odd-cr}(G) \leq \frac{|E(G)|^2}{(2c_2)^2 \log^6 n} + 2|E(G)|n.$$

By Lemma 4.2, there is a partition of the vertex set $V = V_1 \dot\cup V_2$ with $\frac{1}{3} \cdot |V| \leq V_i \leq \frac{2}{3} \cdot |V|$ for $i = 1,2$ and

$$b(G) \leq c_2 \log n \sqrt{\frac{|E(G)|^2}{(2c_2)^2 \log^6 n} + 4n|E(G)|}.$$

If

$$\frac{|E(G)|^2}{(2c_2)^2 \log^6 n} \leq 4n|E(G)|,$$

then we have $|E(G)| \leq c_3 n \log^{c_4 \log k} n$, and we are done. Therefore, we can assume

$$b(G) \leq c_2 \log n \sqrt{\frac{2|E(G)|^2}{(2c_2)^2 \log^6 n}} \leq \frac{|E(G)|}{\log^2 n}.$$

Let $|V_1| = n_1$ and $|V_2| = n_2$. By the induction hypothesis, we have

$$\begin{aligned} |E(G)| &\leq b(G) + c_3 n_1 \log^{c_4 \log k} n_1 + c_3 n_2 \log^{c_4 \log k} n_2 \\ &\leq \tfrac{|E(G)|}{\log^2 n} + c_3 n \log^{c_4 \log k}(2n/3) \\ &\leq \tfrac{|E(G)|}{\log^2 n} + c_3 n (\log n - \log(3/2))^{c_4 \log k}, \end{aligned}$$

which implies

$$|E(G)| \leq c_3 n \log^{c_4 \log k} n \frac{(1 - \log(3/2)/\log n)^{c_4 \log k}}{1 - 1/\log^2 n} \leq c_3 n \log^{c_4 \log k} n. \qquad \square$$

For small values of k, one can obtain better bounds by replacing Lemma 4.1 with a theorem of Pach et al. [13] and Ackerman [1] to obtain

Theorem 4. *For $k > 4$, every n vertex simple topological graph with no $(k,1)$-crossing family has at most $O\left(n \log^{2k+2} n\right)$ edges. For $k = 2,3,4$, every n-vertex simple topological graph with no $(k,1)$-crossing family has at most $O\left(n \log^6 n\right)$ edges.* □

Acknowledgements The authors gratefully acknowledge support from the Swiss National Science Foundation Grant No. 200021-125287/1.

References

1. E. Ackerman, On the maximum number of edges in topological graphs with no four pairwise crossing edges, in *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry (SCG'06)* Sedona, 05–07 June 2006 (ACM, New York, 2006), pp. 259–263
2. P.K. Agarwal, B. Aronov, J. Pach, R. Pollack, M. Sharir, Quasi-planar graphs have a linear number of edges, in *Proceedings of the Symposium on Graph Drawing*, 20–22 September 1995, ed. by F. Brandenburg. Lecture Notes in Computer Science, vol. 1027 (Springer-Verlag, London), pp. 1–7
3. P.K. Agarwal, M. van Kreveld, S. Suri, Label placement by maximum independent set in rectangles. Comput. Geom. Theor. Appl. **11**, 209–218 (1998)
4. E. Ackerman, J. Fox, J. Pach, A. Suk, On grids in topological graphs, in *Proceedings of the 25th ACM Symposium on Computational Geometry (SoCG)*, University of Aarhus, Aarhus, June 2009, pp. 403–412
5. T. Dey, Improved bounds on planar k-sets and k-levels. Discr. Comput. Geom. **19**, 156–161 (1997)
6. J. Fox, J. Pach, Coloring K_k-free intersection graphs of geometric objects in the plane, in *Proceedings of the Twenty-Fourth Annual Symposium on Computational Geometry (SCG'08)*, College Park, MD, 09–11 June 2008 (ACM, New York, 2008), pp. 346–354
7. R. Fulek, J. Pach, A computational approach to Conway's thrackle conjecture. Comput. Geom. **44**(6–7), 345–355 (2011)
   Comput. Geom.: Theor. Appl. (2010), to appear
8. D.S. Hochbaum, W. Maass, Approximation schemes for covering and packing problems in image processing and VLSI. J. ACM **32**, 130–136 (1985)
9. M. Klazar, A. Marcus, Extensions of the linear bound in the Füredi–Hajnal conjecture. Adv. Appl. Math. **38**, 258–266 (2006)
10. L. Lovász, J. Pach, M. Szegedy, On Conway's thrackle conjecture. Discr. Comput. Geom. **18**(4), 369–376 (1997)
11. J. Matoušek, *Lectures on Discrete Geometry*, Springer-Verlag, New York (2002)
12. A. Marcus, G. Tardos, Excluded permutation matrices and the Stanley–Wilf conjecture. J. Comb. Theor. Ser. A **107**(1), 153–160 (2004)
13. J. Pach, R. Radoicic, G. Toth, Relaxing planarity for topological graphs, in *Discrete and Computational Geometry, Japanese Conference, JCDCG 2002*, Tokyo, Japan, ed. by J. Akiyama, M. Kano. Lecture Notes in Computer Science, vol. 2866 (Springer, 2003), pp. 221–232
14. J. Pach, F. Shahrokhi, M. Szegedy, Applications of the crossing number, in *Proceedings of the Tenth Annual Symposium on Computational Geometry (SCG'94)*, Stony Brook, NY, 06–08 June 1994 (ACM, New York, 1994), pp. 198–202
15. J. Pach, G. Tóth, Disjoint edges in topological graphs, in *Combinatorial Geometry and Graph Theory*, ed. by J. Akiyama et al., Lecture Notes in Computer Science, vol. 3330 (Springer, Berlin, 2005), pp. 133–140
16. J. Pach, G. Tóth, Which crossing number is it anyway? J. Comb. Theor. Ser. B **80**(2), 225–246 (2000)
17. J. Pach, J. Töröcsik, Some geometric applications of Dilworth's theorem. Discr. Comput. Geom. **12**, 1–7 (1994)

18. G. Tardos, G. Tóth, Crossing stars in topological graphs, in *Japan Conference on Discrete and Computational Geometry*. Lecture Notes in Computer Science, vol. 3742 (Springer, Berlin, 2004), pp. 184–197
19. G. Tóth, Note on geometric graphs. J. Comb. Theor. Ser. A **89**(1), 126–132 (2000)
20. G. Tóth, P. Valtr, Geometric graphs with few disjoint edges. Technical report. UMI Order Number: 98-22. Center for Discrete Mathematics & Theoretical Computer Science, 1998
21. P. Valtr, Graph drawings with no k pairwise crossing edges, in *Graph Drawing '97*, Rome, September 1997, ed. by G. Di Battista, pp. 205–218

Counting Plane Graphs: Flippability and Its Applications

Michael Hoffmann, André Schulz, Micha Sharir, Adam Sheffer, Csaba D. Tóth, and Emo Welzl

Abstract We generalize the notions of flippable and simultaneously flippable edges in a triangulation of a set S of points in the plane to *pseudo-simultaneously flippable edges*. Such edges are related to the notion of convex decompositions spanned by S.

We prove a worst-case tight lower bound for the number of pseudo-simultaneously flippable edges in a triangulation in terms of the number of vertices. We use this bound for deriving new upper bounds for the maximal number of crossing-free straight-edge graphs that can be embedded on any fixed set of N points in the plane. We obtain new upper bounds for the number of spanning trees and forests as well. Specifically, let $\mathrm{tr}(N)$ denote the maximum number of triangulations on a set of N points in the plane. Then we show [using the known bound $\mathrm{tr}(N) < 30^N$] that any N-element point set admits at most $6.9283^N \cdot \mathrm{tr}(N) < 207.85^N$ crossing-free straight-edge graphs, $O(4.7022^N) \cdot \mathrm{tr}(N) = O(141.07^N)$ spanning

M. Hoffmann • E. Welzl
Institute of Theoretical Computer Science, ETH Zürich, 8092 Zürich, Switzerland
e-mail: hoffmann@inf.ethz.ch; welzl@inf.ethz.ch

A. Schulz
Institut für Mathematische Logik und Grundlagenforschung,
Universität Münster, Münster, Germany
e-mail: andre.schulz@uni-muenster.de

M. Sharir
School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel

Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA
e-mail: michas@tau.ac.il

A. Sheffer (✉)
School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel
e-mail: sheffera@tau.ac.il

C.D. Tóth
Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada
e-mail: cdtoth@ucalgary.ca

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_16,

trees, and $O(5.3514^N) \cdot \mathsf{tr}(N) = O(160.55^N)$ forests. We also obtain upper bounds for the number of crossing-free straight-edge graphs that have cN, fewer than cN, or more than cN edges, for any constant parameter c, in terms of c and N.

1 Introduction

A *crossing-free straight-edge graph* G is an embedding of a planar graph in the plane such that the vertices are mapped to a set S of points in the plane and the edges are pairwise noncrossing line segments between pairs of points in S. (Segments are allowed to share endpoints.) By Fáry's classical result [10], such an embedding is always possible. In this chapter, we fix a labeled set S of points in the plane, and we only consider planar graphs that admit a straight-edge embedding with vertex set S. By *labeled*, we mean that each vertex of the graph has to be mapped to a unique designated point of S. Analysis of the number of plane embeddings of planar graphs in which the set of vertices is not restricted to a specific embedding, or when the vertices are not labeled, can be found, for example, in [16, 21, 32].

A *triangulation* of a set S of N points in the plane is a maximal crossing-free straight-edge graph on S (that is, no additional straight edges can be inserted without crossing some of the existing edges). Triangulations are an important geometric construct that are used in many algorithmic applications and are also an interesting object of study in discrete and combinatorial geometry (recent comprehensive surveys can be found in [7, 17]).

Improving the bound on the maximum number of triangulations that any set of N points in the plane can have has been a major research theme during the past 30 years. The initial upper bound 10^{13N} of [2] has been steadily improved in several papers (e.g., see [8,25,29]), culminating with the current record of 30^N due to Sharir and Sheffer [26]. Other papers have studied lower bounds on the maximal number of triangulations (e.g., [1, 9]), and upper or lower bounds on the number of other kinds of planar graphs (e.g., [5, 6, 23, 24]).

Every triangulation of S contains the edges of the convex hull of S, and the remaining edges of the triangulation decompose the interior of the convex hull into triangular faces. Assume that S contains N points, h of which are on the convex hull boundary and the remaining $n = N - h$ points are interior to the hull (we use this notation throughout). By Euler's formula, every triangulation of S has $3n + 2h - 3$ edges (h *hull* edges, common to all triangulations, and $3n + h - 3$ *interior* edges, each adjacent to two triangles), and $2n + h - 2$ bounded triangular faces.

Edge Flips. Edge flips are simple operations that replace one or several edges of a triangulation with new edges and produce a new triangulation. As we will see in Sect. 3, edge flips are instrumental for counting various classes of subgraphs in triangulations. In the next few paragraphs, we review previous results on edge flips, and propose a new type of edge flip. We say that an interior edge in a triangulation of S is *flippable* if its two adjacent triangles form a convex quadrilateral. A flippable edge

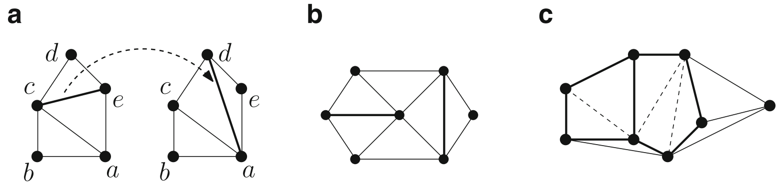


Fig. 1 (**a**) The edge *ce* can be flipped to the edge *ad*. (**b**) The two bold edges are simultaneously flippable. (**c**) Interior-disjoint convex quadrilateral and convex pentagon in a triangulation.

can be *flipped*, that is, removed from the graph of the triangulation and replaced by the other diagonal of the corresponding quadrilateral, thereby obtaining a new triangulation of S. An edge-flip operation is depicted in Fig. 1a, where the edge *ce* is flipped to the edge *ad*. Already in 1936, Wagner [34] has shown that any *unlabeled abstract* triangulation T (in this case, two triangulations are considered identical if we can relabel and change the planar embedding of the vertices of the first triangulation, to obtain the second triangulation) can be transformed into any other triangulation T' (with the same number of vertices) through a series of edge flips (here one uses a more abstract notion of an edge flip). When we deal with a pair of triangulations over a specific common (labeled) set S of points in the plane, there always exists such a sequence of $O(|S|^2)$ flips, and this bound is tight in the worst case (e.g., see [3, 19]). Moreover, there are algorithms that perform such sequences of flips to obtain some "optimal" triangulation (typically, the Delaunay triangulation; see [12] for example), which, as a byproduct, provide an edge-flip sequence between any specified pair of triangulations of S.

How many flippable edges can a single triangulation have? Given a triangulation T, we denote by $\mathsf{flip}(T)$ the number of flippable edges in T. Hurtado, Noy, and Urrutia [19] proved the following lower bound.

Lemma 1.3 ([19]). *For any triangulation T over a set of N points in the plane,*

$$\mathsf{flip}(T) \geq N/2 - 2.$$

Moreover, there are triangulations (of specific point sets of arbitrarily large size) for which this bound is tight.

To obtain a triangulation with exactly $N/2 - 2$ flippable edges (for an even N), start with a convex polygon with $N/2 + 1$ vertices, triangulate it in some arbitrary manner, insert a new point into each of the $N/2 - 1$ resulting bounded triangles, and connect each new point p to the three hull vertices that form the triangle containing p. Such a construction is depicted in Fig. 2. The resulting graph is a triangulation with N vertices and exactly $N/2 - 2$ flippable edges, namely, the chords of the initial triangulation.

Next, we say that two flippable edges e and e' of a triangulation T are *simultaneously flippable* if no triangle of T is incident to both edges; equivalently,

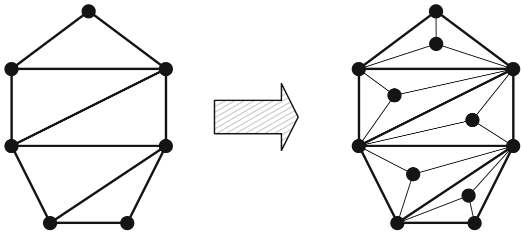


Fig. 2 Constructing a triangulation with $N/2-2$ flippable edges

the quadrilaterals corresponding to e and e' are interior-disjoint. See Fig. 1b for an illustration. Notice that flipping an edge e cannot affect the flippability of any edge simultaneously flippable with e. Given a triangulation T, let $\mathsf{flip}_s(T)$ denote the size of the largest subset of edges of T such that every pair of edges in the subset is simultaneously flippable. The following lemma, improving upon an earlier weaker bound in [13], is taken from Souvaine et al. [30].

Lemma 1.4 ([30]). *For any triangulation T over a set of N points in the plane,* $\mathsf{flip}_s(T) \geq (N-4)/5$.

Galtier et al. [13] show that this bound is tight in the worst case, by presenting a specific triangulation in which at most $(N-4)/5$ edges are simultaneously flippable.

Pseudo-Simultaneously Flippable Edge Sets. A set of simultaneously flippable edges in a triangulation T can be considered the set of diagonals of a collection of interior-disjoint convex quadrilaterals. We consider a more liberal definition of simultaneously flippable edges, by taking, within a fixed triangulation T, the diagonals of a set of interior-disjoint convex polygons, each with at least four edges (so that the boundary edges of these polygons belong to T). Consider such a collection of convex polygons $Q_1,\ldots,Q_m$, where Q_i has $k_i \geq 4$ edges, for $i=1,\ldots,m$. We can then retriangulate each Q_i independently, to obtain many different triangulations. Specifically, each Q_i can be triangulated in C_{k_i-2} ways, where C_j is the jth *Catalan number* (see, e.g., [31, Sect. 5.3]). Hence, we can get $M=\prod_{i=1}^{m} C_{k_i-2}$ different triangulations in this way. In particular, if a graph $G \subseteq T$ (namely, all the edges of G are edges of T) does not contain any diagonal of any Q_i (it may contain boundary edges though) then G is a subgraph of (at least) M distinct triangulations. An example is depicted in Fig. 1c, where by "flipping" (or rather, redrawing) the diagonals of the highlighted quadrilateral and pentagon, we can get $C_2 \cdot C_3 = 2 \cdot 5 = 10$ different triangulations (including the one shown), and any subgraph of the triangulation that does not contain any of these diagonals is a subgraph of these 10 triangulations. We say that a set of interior edges in a triangulation is *pseudo-simultaneously flippable* (*ps-flippable* for short) if, after the deletion of these edges, every bounded face of the remaining graph is convex, and there are no vertices of degree 0. Notice that all three notions of flippability are defined within a fixed triangulation T of S (although each of them gives a recipe for producing many other triangulations).

Table 1 Bounds for minimum numbers of the various types of flippable edges in a triangulation of N points

Edge type	Lower bound
Flippable	$N/2-2$ [19]
Simultaneously flippable	$N/5-4/5$ [13, 30]
ps-Flippable	$\max\{N/2-2, h-3\}$

All of these bounds are tight in the worst case

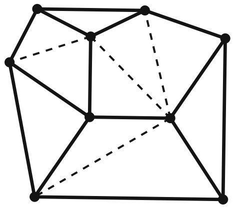


Fig. 3 A convex decomposition of S. When completing it into a triangulation, the added (*dashed*) diagonals form a set of ps-flippable edges. This is one of the $C_2 \cdot C_2 \cdot C_3 = 20$ possible completions

Our Results. In Sect. 2, we derive a lower bound on the size of the largest set of ps-flippable edges in a triangulation, and show that this bound is tight in the worst case. Specifically, we have the following result.

Lemma 1.5 (ps-Flippability lemma). *Let S be a set of N points in the plane, and let T be a triangulation of S. Then T contains a set of at least* $\max\{N/2-2, h-3\}$ *ps-flippable edges. This bound is tight in the worst case.*

Table 1 summarizes the bounds for the minimum numbers of the various types of flippable edges in a triangulation.

We also relate ps-flippable edges to *convex decompositions* of S. These are crossing-free straight-edge graphs on S such that (a) they include all the hull edges, (b) each of their bounded faces is a convex polygon, and (c) no point of S is isolated. See Fig. 3 for an illustration.

Counting Plane Graphs: New Upper Bounds. In Sect. 3, we use Lemma 1.5 to derive several upper bounds on the numbers of planar graphs of various kinds embedded as crossing-free straight-edge graphs on a fixed labeled set S. For a set S of points in the plane, we denote by $\mathcal{T}(S)$ the set of all triangulations of S, and put $\mathsf{tr}(S) := |\mathcal{T}(S)|$. Similarly, we denote by $\mathcal{P}(S)$ the set of all crossing-free straight-edge graphs on S, and put $\mathsf{pg}(S) := |\mathcal{P}(S)|$. We also let $\mathsf{tr}(N)$ and $\mathsf{pg}(N)$ denote, respectively, the maximum values of $\mathsf{tr}(S)$ and of $\mathsf{pg}(S)$, over all sets S of N points in the plane.

Since a triangulation of S has fewer than $3|S|$ edges, the trivial upper bound $\mathsf{pg}(S) < 8^{|S|} \cdot \mathsf{tr}(S)$ holds for any point set S. Recently, Razen, Snoeyink, and Welzl [22] slightly improved the upper bound on the ratio $\mathsf{pg}(S)/\mathsf{tr}(S)$ from $8^{|S|}$ down to $O\left(7.9792^{|S|}\right)$. We give a more significant improvement on the ratio with an upper bound of $6.9283^{|S|}$. Combining this bound with the recent bound $\mathsf{tr}(S) < 30^{|S|}$ [26], we get $\mathsf{pg}(N) < 207.85^N$. We provide similar improved ratios and absolute bounds

Table 2 Upper and lower bounds for the number of several types of crossing-free straight-edge graphs on a set of N points in the plane

Graph type	Lower bound	Previous upper bound	New upper bound	In the form $a^N \cdot \mathrm{tr}(N)$
Plane graphs	$\Omega(41.18^N)$ [1]	$O(239.40^N)$ [22, 26]	$\mathbf{207.85^N}$	$\mathbf{6.9283^N \cdot tr(N)}$
Spanning trees	$\Omega(11.97^N)$ [9]	$O(158.56^N)$ [6, 26]	$\mathbf{O(141.07^N)}$	$\mathbf{O(4.7022^N) \cdot tr(N)}$
Forests	$\Omega(12.23^N)$ [9]	$O(194.66^N)$ [6, 26]	$\mathbf{O(160.55^N)}$	$\mathbf{O(5.3514^N) \cdot tr(N)}$

By plane graphs, we mean all crossing-free straight-edge graphs embedded on a specific labeled point set. Our new bounds are in the right two columns.

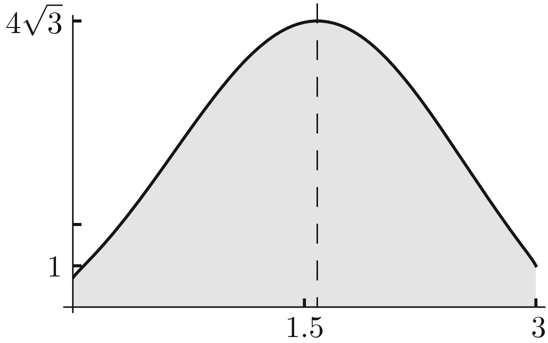


Fig. 4 The base $B(c)$ of the exponential factor in the bound on the number of crossing-free straight-edge graphs with $c|S|$ edges, as a function of c. The maximum is attained at $c = 19/12$ (see below)

for the numbers of crossing-free straight-edge spanning trees and forests (i.e., cycle-free graphs). Table 2 summarizes these results. [1]

We also derive similar ratios for the number of crossing-free straight-edge graphs with exactly $c|S|$ edges, with at least $c|S|$ edges, and with at most $c|S|$ edges, for $0 < c < 3$. For the case of crossing-free straight-edge graphs with exactly $c|S|$, we obtain the bound[2]

$$O^*\left(\left(\frac{5^{5/2}}{8(c+t-1/2)^{c+t-1/2}(3-c-t)^{3-c-t}(2t)^t(1/2-t)^{1/2-t}}\right)^N \cdot \mathrm{tr}(S)\right),$$

where $t = \frac{1}{2}\left(\sqrt{(7/2)^2+3c+c^2} - 5/2 - c\right)$. Figure 4 contains a plot of the base $B(c)$ of the exponential factor multiplying $\mathrm{tr}(N)$ in this bound, as a function of c.

Notation. Here are some additional notations that we use.

[1] Up-to-date bounds for these and for other families of graphs can be found in http://www.cs.tau.ac.il/~sheffera/counting/PlaneGraphs.html (version of November 2010).

[2] In the notations $O^*()$, $\Theta^*()$, and $\Omega^*()$, we neglect polynomial factors.

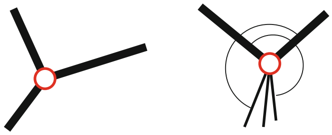


Fig. 5 Separable edges

For a triangulation T and an integer $i \geq 3$, let $v_i(T)$ denote the number of interior vertices of degree i in T.

Given two crossing-free straight-edge graphs G and H over the same point set S, we write $G \subseteq H$ to indicate that every edge in G is also an edge in H.

Similarly to the case of edges, the *hull vertices* (respectively, *interior vertices*) of a set S of points in the plane are those that are part of the boundary of the convex hull of S (respectively, not part of the convex hull boundary).

We only consider point sets S in general position; that is, no three points in S are collinear. For upper bounds on the number of graphs, this involves no loss of generality, because the number of graphs can only increase if collinear points are slightly perturbed into general position.

Separable Edges. Let p be an interior vertex in a convex decomposition G of S. Following the notation in [27], we call an edge e incident to p in G *separable at* p if it can be separated from the other edges incident to p by a line through p (see Fig. 5, where the separating lines are not drawn). Equivalently, edge e is separable at p if the two angles between e and its clockwise and counterclockwise neighboring edges (around p) sum up to more than π. Following [19], we observe the following easy properties, both of which materialize in Fig. 5.

(i) If p is an interior vertex of degree 3 in G, its three incident edges are separable at p, for otherwise p would have been a reflex vertex of some face.
(ii) An interior vertex p of degree 4 or higher can have at most two incident edges that are separable at p (and if it has two such edges, they must be consecutive in the circular order around p).

2 The Size of ps-Flippable Edge Sets

In this section, we establish the ps-flippability lemma (Lemma 1.5 from the introduction). We restate the lemma for the convenience of the reader.

Lemma 1.5 (ps-Flippability lemma). *Let S be a set of N points in the plane, and let T be a triangulation of S. Then T contains a set of at least* $\max\{N/2-2, h-3\}$ *ps-flippable edges. This bound is tight in the worst case.*

Proof. Starting with the proof of the lower bound, we apply the following iterative process to T. As long as there exists an interior edge whose removal does not create a nonconvex face, we pick such an edge and remove it. When we stop, we have a

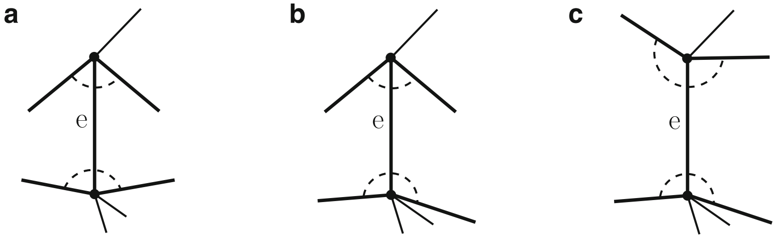


Fig. 6 (**a**) An edge not separable at both of its endpoints can be removed from the graph. (**b**, **c**) An edge separable in at least one of its endpoints cannot be removed from the convex decomposition

crossing-free straight-edge graph G, all of whose bounded faces are convex; that is, we have a locally minimal convex decomposition of S. Note that all h original hull edges are still in G and that every interior vertex of G has degree at least 3 (recall the general position assumption).

Note that every edge of G is separable at one or both of its endpoints, for we can remove any other edge and the graph will continue to have only convex faces (see Fig. 6). We denote by m the number of edges of G, and by m_{int} the number of its interior edges. Recalling properties (i) and (ii) of separable edges, we have $m = m_{\text{int}} + h$ and

$$m_{\text{int}} \leq 3v_3 + 2v_{4,2} + v_{4,1} =: a, \tag{1}$$

where v_3 is the number of interior vertices of degree 3 in G, and $v_{4,i}$ is the number of interior vertices u of degree at least 4 in G with exactly i edges separable at u. Notice that

$$n = v_3 + v_{4,0} + v_{4,1} + v_{4,2}.$$

The estimate in (1) may be pessimistic, because it doubly counts edges that are separable at both endpoints (such as the one in Fig. 6c). To address this possible overestimation, denote by m_{double} the number of edges that are separable at both endpoints, which we refer to as *doubly separable edges*, and rewrite (1) as

$$m_{\text{int}} = 3v_3 + 2v_{4,2} + v_{4,1} - m_{\text{double}} = a - m_{\text{double}}. \tag{2}$$

Denoting by f the number of bounded faces of G, we have, by Euler's formula,

$$n + h + (f + 1) = (m_{\text{int}} + h) + 2$$

(the expression in the parentheses on the left is the number of faces in G, and the expression in the parentheses on the right is the number of edges), or

$$f = m_{\text{int}} - n + 1. \tag{3}$$

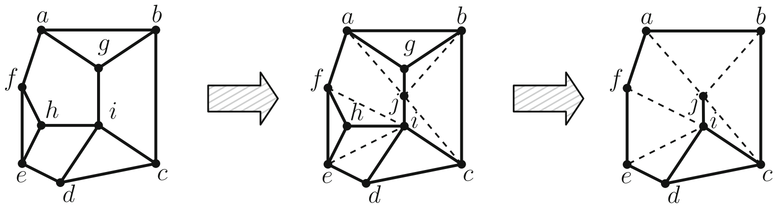


Fig. 7 A convex decomposition G of S, its corresponding graph G', and the reduced form of G' after removing vertices of degree 3. The edges that have been added are *dashed*

Let f_k, for $k \geq 3$, denote the number of interior faces of degree k in G. By doubly counting the number of edges in G, and then applying (3), we get

$$\sum_{k\geq 3} kf_k = 2m_{\text{int}} + h = 2(f+n-1)+h = \sum_{k\geq 3} 2f_k + 2n - 2 + h$$

or

$$\sum_{k\geq 3}(k-2)f_k = 2n+h-2. \tag{4}$$

The number of edges that were removed from T is $\sum_{k\geq 3}(k-3)f_k$, because a face of G of degree k must have had $k-3$ diagonals that were edges of T. This number is therefore

$$\sum_{k\geq 3}(k-3)f_k = \sum_{k\geq 3}(k-2)f_k - f = 2n+h-2-f =$$
$$= 2n+h-2-(m_{\text{int}}-n+1) = 3n+h-3-a+m_{\text{double}} \tag{5}$$

[by first applying (4), then (3), and finally (2)].

We next derive a lower bound for the right-hand side of (5). For this, we transform G into another graph G' as follows. We first subdivide each doubly separable edge of G at its midpoint, say, and add the subdivision point as a new vertex of G (e.g., see the vertex j in Fig. 7). We now modify G as follows. We take each vertex u of degree 3 in G and surround it by a triangle, by connecting all pairs of its three neighbors. Notice that some of these neighbors may be new subdivision vertices and that some of the edges of the surrounding triangle may already belong to G. For example, see Fig. 7, where the edges ei, fi are added around the vertex h and the edges aj, bj are added around the vertex g. Next we take each interior vertex u with two separable edges at u and complete these two edges into a triangle by connecting their other endpoints, each of which is either an original point or a new subdivision point; here too the completing edge may already belong to G. For example, see the edge cj in Fig. 7, induced by the two separable edges of the vertex i. We then take the resulting graph G' and remove each vertex of degree 3 and its three incident edges; see the reduced version of G' in Fig. 7. A crucial and easily verified property of this transformation is that the newly embedded edges do not cross each other, nor do they cross old edges of G.

The number f' of bounded faces of the new graph G' is at least $v_3 + v_{4,2}$, which is the number of triangles that we have created, and the number n' of its interior vertices is $n - v_3 + m_{\text{double}}$. Also, G' still has h hull edges. Using Euler's formula, as in (3) and (4) above, we have $f' \le 2n' + h - 2$. Combining the above, we get

$$v_3 + v_{4,2} \le 2(n - v_3 + m_{\text{double}}) + h - 2$$

or

$$m_{\text{double}} \ge \frac{3}{2}v_3 + \frac{1}{2}v_{4,2} - n - \frac{1}{2}h + 1.$$

Hence, the right-hand side of (5) is at least
$3n + h - 3 - a + m_{\text{double}}$

$$\begin{aligned}
&\ge 3n + h - 3 - (3v_3 + 2v_{4,2} + v_{4,1}) + \left(\frac{3}{2}v_3 + \frac{1}{2}v_{4,2} - n - \frac{1}{2}h + 1\right) \\
&= 2n + \frac{1}{2}h - 2 - \frac{3}{2}v_3 - \frac{3}{2}v_{4,2} - v_{4,1} \\
&\ge 2n + \frac{1}{2}h - 2 - \frac{3}{2}(v_3 + v_{4,2} + v_{4,1} + v_{4,0}) \\
&= \frac{n+h}{2} - 2 = \frac{N}{2} - 2.
\end{aligned}$$

In other words, the number of edges that we have removed from T is at least $N/2 - 2$. On the other hand, we always have $m_{\text{double}} \ge 0$ and $a \le 3n$. Substituting these trivial bounds in (5), we get at least $h - 3$ ps-flippable edges. This completes the proof of the lower bound.

It is easily noticed that only flippable edges of T could have been removed in the initial pruning stage. Hurtado, Noy, and Urrutia [19] present two distinct triangulations that contain exactly $N/2 - 2$ flippable edges (one of those is depicted in Fig. 2). These triangulations cannot have a set of more than $N/2 - 2$ ps-flippable edges. Therefore, there are triangulations for which our bound is tight in the worst case. Similarly, for point sets in convex position, all $h - 3$ interior edges form a set of ps-flippable edges, showing that the other term in the lower bound is also tight in the worst case. □

Remark. The proof of Lemma 1.5 actually yields the slightly better bound

$$\frac{1}{2}N + \frac{1}{2}v_{4,1} + \frac{3}{2}v_{4,0} - 2.$$

That is, for the bound to be tight, every interior vertex u of degree 4 or higher must have two incident edges separable at u (note that this condition holds vacuously for the triangulation in Fig. 2).

Convex Decompositions. The preceding analysis is also related to the notion of *convex decompositions*, as defined in the Introduction. Urrutia [33] asked what is the minimum number of faces that can always be achieved in a convex decomposition of any set of N points in the plane? Hosono [18] proved that every planar set of N points admits a convex decomposition with at most $\lceil \frac{7}{5}(N+2) \rceil$ (bounded) faces. For every $N \geq 4$, García-Lopez and Nicolás [15] constructed N-element point sets that do not admit a convex decomposition with fewer than $\frac{12}{11}N-2$ faces. By Euler's formula, if a connected crossing-free straight-edge graph has N vertices and e edges, then it has $e-N+2$ faces (including the exterior face). It follows that for convex decompositions, minimizing the number of faces is equivalent to minimizing the number of edges. (For convex decompositions contained in a given triangulation, this is also equivalent to maximizing the number of removed edges, which form a set of ps-flippable edges.)

Lemma 1.5 directly implies the following corollary. (The bound that it gives is weaker than the bound in [18], but it holds for every triangulation.)

Corollary 2.1. *Let S be a set of N points in the plane, so that its convex hull has h vertices, and let T be a triangulation of S. Then T contains a convex decomposition of S with at most $\frac{3}{2}N-h \leq \frac{3}{2}N-3$ convex faces and at most $\frac{5}{2}N-h-1 \leq \frac{5}{2}N-4$ edges. Moreover, there exist point sets S of arbitrarily large size, and triangulations $T \in \mathcal{T}(S)$ for which these bounds are tight.*

3 Applications of ps-Flippable Edges to Graph Counting

In this section/ we apply the ps-flippability lemma (Lemma 1.5) to obtain several improved bounds on the number of crossing-free straight-edge graphs of various kinds on a fixed set of points in the plane.

3.1 The Ratio Between the Number of Crossing-Free Straight-Edge Graphs and the Number of Triangulations

We begin by recalling some observations already made in the Introduction. Let S be a set of N points in the plane. Every crossing-free straight-edge graph in $\mathcal{P}(S)$ is contained in at least one triangulation in $\mathcal{T}(S)$. Additionally, since a triangulation has fewer than $3N$ edges, every triangulation $T \in \mathcal{T}(S)$ contains fewer than $2^{3N} = 8^N$ crossing-free straight-edge graphs. This immediately implies

$$\mathsf{pg}(S) < 8^N \cdot \mathsf{tr}(S).$$

However, this inequality seems rather weak since it potentially counts some crossing-free straight-edge graphs many times. More formally, given a graph $G \in \mathcal{P}(S)$ contained in x distinct triangulations of S, we say that G has a *support* of x, and write $\mathsf{supp}(G) = x$. Thus, every graph $G \in \mathcal{P}(S)$ will be counted $\mathsf{supp}(G)$ times in the preceding inequality.

Recently, Razen et al. [22] managed to break the 8^N barrier by overcoming the above inefficiency. However, they obtained only a slight improvement, with the bound $\mathsf{pg}(S) = O\left(7.9792^N\right) \cdot \mathsf{tr}(S)$. We now present a more significant improvement, using a much simpler technique that relies on the ps-flippability lemma.

Theorem 1. *For every set S of N points in the plane, h of which are on the convex hull,*

$$\mathsf{pg}(S) \leq \begin{cases} \dfrac{(4\sqrt{3})^N}{2^h} \cdot \mathsf{tr}(S) < \dfrac{6.9283^N}{2^h} \cdot \mathsf{tr}(S), & \text{for } h \leq N/2, \\ 8^N (3/8)^h \cdot \mathsf{tr}(S), & \text{for } h > N/2. \end{cases}$$

Proof. The exact value of $\mathsf{pg}(S)$ is easily seen to be

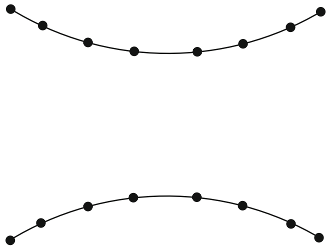
$$\mathsf{pg}(S) = \sum_{T \in \mathcal{T}(S)} \sum_{\substack{G \in \mathcal{P}(S) \\ G \subseteq T}} \frac{1}{\mathsf{supp}(G)}, \tag{6}$$

because every graph G appears $\mathsf{supp}(G)$ times in the sum, and thus contributes a total of $\mathsf{supp}(G) \cdot \frac{1}{\mathsf{supp}(G)} = 1$ to the count. We obtain an upper bound on this sum as follows. Consider a graph $G \in \mathcal{P}(S)$ and a triangulation $T \in \mathcal{T}(S)$, such that $G \subseteq T$. By Lemma 1.5, there is a set F of $t = \max(N/2-2, h-3)$ ps-flippable edges in T.[3] Let $F_{\bar{G}}$ denote the set of edges that are in F but *not* in G, and put $j = |F_{\bar{G}}|$. Removing the edges of $F_{\bar{G}}$ from T yields a convex decomposition of S that still contains G and whose nontriangular interior faces have a total of j missing diagonals. Suppose that there are m such faces, with $j_1, j_2, \ldots, j_m$ diagonals, respectively, where $\sum_{k=1}^{m} j_k = j$. Then these faces can be triangulated in $\prod_{k=1}^{m} C_{j_k+1}$ ways, and each of the resulting triangulations contains G. We always have $C_{i+1} \geq 2^i$, for any $i \geq 1$, as is easily verified, and so $\mathsf{supp}(G) \geq 2^j$. (Equality occurs when all the nontriangular faces of $T \setminus F_{\bar{G}}$ are quadrilaterals.)

Next, we estimate the number of subgraphs $G \subseteq T$ for which the set $F_{\bar{G}}$ is of size j. Denote by E the set of edges of T that are not in F, and assume that the convex hull of S has h vertices. Since there are $3N-3-h$ edges in any triangulation of S, $|E| \leq 3N-3-h-t$. To obtain a graph G for which $|F_{\bar{G}}| = j$, we choose any subset of edges from E, and any j edges from F (the j edges of F that will not belong to G). Therefore, the number of such subgraphs is at most $2^{3N-h-t-3} \cdot \binom{t}{j}$.

[3] Here we implicitly assume that N is even. The case where N is odd is handled in the exact same manner, since a constant change in the size of F does not affect the asymptotic bounds.

Fig. 8 A double-chain configuration with 16 vertices



We can thus rewrite (6) to obtain

$$\begin{aligned}\mathsf{pg}(S) &\leq \sum_{T\in\mathcal{T}(S)}\sum_{j=0}^{t} 2^{3N-h-t-3}\cdot\binom{t}{j}\cdot\frac{1}{2^j}\\ &= \mathsf{tr}(S)\cdot 2^{3N-h-t-3}\sum_{j=0}^{t}\binom{t}{j}\frac{1}{2^j}\\ &= \mathsf{tr}(S)\cdot 2^{3N-h-t-3}\cdot(3/2)^t.\end{aligned}$$

If $t = N/2-2$, we get $\mathsf{pg}(S) < \mathsf{tr}(S)\cdot\frac{(4\sqrt{3})^N}{2^h} < \frac{6.9283^N}{2^h}\cdot\mathsf{tr}(S)$. If $t = h-3$, we have $\mathsf{pg}(S)\leq\mathsf{tr}(S)\cdot 2^{3N-2h}\cdot(3/2)^h = \mathsf{tr}(S)\cdot 8^N\cdot(3/8)^h$. To complete the proof, we note that $N/2-2 > h-3$ when $h\leq N/2$. □

For a lower bound on $\mathsf{pg}(S)/\mathsf{tr}(S)$, we consider the *double-chain* configurations, presented in [14] (and depicted in Fig. 8). It is shown in [14] that when S is a double-chain configuration, $\mathsf{tr}(S) = \Theta^*\left(8^N\right)$ and $\mathsf{pg}(S) = \Theta^*\left(39.8^N\right)$ [actually, only the lower bound on $\mathsf{pg}(N)$ is given in [14]; the upper bound appears in [1]]. Thus, we have $\mathsf{pg}(S) = \Theta^*\left(4.975^N\right)\cdot\mathsf{tr}(S)$ (for this set $h = 4$, so h has no real effect on the asymptotic bound of Theorem 1).

For another lower bound, consider the case where S is in convex position. In this case, we have $\mathsf{tr}(S) = C_{N-2} = \Theta^*(4^N)$ and $\mathsf{pg}(S) = \Theta^*(11.65^N)$ (see [11]). Hence, $\mathsf{pg}(S)/\mathsf{tr}(S) = \Theta^*(2.9125^N)$, whereas the upper bound provided by Theorem 1 is 3^N in this case. Informally, the (rather small) discrepancy between the exact bound in [11] and our bound in the convex case comes from the fact that when j is large, the faces of the resulting convex decomposition are likely to have many edges, which makes $\mathsf{supp}(G)$ substantially larger than 2^j. It is an interesting open problem to exploit this observation to improve our upper bound when h is large.

Finally, recall the notations $\mathsf{tr}(N) = \max_{|S|=N}\mathsf{tr}(S)$ and $\mathsf{pg}(N) = \max_{|S|=N}\mathsf{pg}(S)$. Combining the bound $\mathsf{tr}(N) < 30^N$, obtained in [26], with the first bound of Theorem 1 implies $\mathsf{pg}(N) < 207.85^N$; see Table 2 for comparison with earlier bounds. The bound improves significantly as h gets larger.

3.2 The Number of Spanning Trees and Forests

Spanning Trees. For a set S of N points in the plane, we denote by $\mathcal{ST}(S)$ the set of all crossing-free straight-edge spanning trees of S, and put $\mathsf{st}(S) := |\mathcal{ST}(S)|$. Moreover, we let $\mathsf{st}(N) = \max_{|S|=N} \mathsf{st}(S)$.

Buchin and Schulz [6] have recently shown that every crossing-free straight-edge graph contains $O\left(5.2852^N\right)$ spanning trees, improving upon the earlier bound of $5.\bar{3}^N$ due to Ribó Mor and Rote [23, 24]. We thus get $\mathsf{st}(S) = O\left(5.2852^N\right) \cdot \mathsf{tr}(S)$ for every set S of N points in the plane. The bound from [6] cannot be improved much further, since there are triangulations with at least 5.0295^N spanning trees [23, 24]. However, the ratio between $\mathsf{st}(S)$ and $\mathsf{tr}(S)$ can be improved beyond that bound, by exploiting and overcoming the same inefficiency as in the case of all crossing-free straight-edge graphs; that is, the fact that some spanning trees may get multiply counted in many triangulations.

We now derive such an improved ratio by using ps-flippable edges. The proof goes along the same lines of the proof of Theorem 1.

Theorem 2. *For every set S of N points in the plane,*

$$\mathsf{st}(S) = O\left(4.7022^N\right) \cdot \mathsf{tr}(S).$$

Proof. The exact value of $\mathsf{st}(S)$ is

$$\mathsf{st}(S) = \sum_{T \in \mathcal{T}(S)} \sum_{\substack{\tau \in \mathcal{ST}(S) \\ \tau \subset T}} \frac{1}{\mathsf{supp}(\tau)}.$$

Consider a spanning tree $\tau \in \mathcal{ST}(S)$ and a triangulation $T \in \mathcal{T}(S)$, such that $\tau \subset T$. As in Theorem 1, let F be a set of $N/2-2$ ps-flippable edges in T. (Here we do not exploit the alternative bound of $h-3$ on the size of F.) Also, let $F_{\bar{\tau}}$ denote the set of edges that are in F but *not* in τ, and put $j = |F_{\bar{\tau}}|$. Thus, as argued earlier, $\mathsf{supp}(\tau) \geq 2^j$.

Next, we estimate the number of spanning trees $\tau \subset T$ for which the set $F_{\bar{\tau}}$ is of size j. First, there are $\binom{|F|}{j} < \binom{N/2}{j}$ ways to choose the j edges of F that τ does not use. We next contract the $N/2-2-j$ edges of F that were chosen to be in τ (which will result in having some parallel edges, and possibly also loops) and then remove the remaining edges of F. This produces a nonsimple graph G with $N/2+2+j$ vertices and fewer than $5N/2$ edges (recall that by Euler's formula, G contains at most $3N-6$ edges). Let S' denote the set of vertices of G, and let d_v denote the degree in G of a point $v \in S$. As shown in [6, 23], the number of spanning trees in a graph G (not necessarily planar or simple) is at most the product of the vertex degrees in G. Thus, the number of ways to complete the tree is at most

$$\prod_{v \in S'} d_v \leq \left(\frac{\sum_{v \in S'} d_v}{|S'|}\right)^{|S'|} < \left(\frac{5N}{N/2+2+j}\right)^{N/2+2+j}$$

(where we have used the inequality of means for the first inequality). Hence, there are fewer than $\binom{N/2}{j} \cdot \left(\frac{5N}{N/2+2+j}\right)^{N/2+2+j}$ spanning trees $\tau \subset T$ with $|F_{\bar{\tau}}| = j$. However, when j is large, it is better to use the bound $O\left(5.2852^N\right)$ from [6] instead.[4]

We thus get, for a threshold parameter $a < 0.25$ that we will set in a moment,

$$\mathsf{st}(S) < \sum_{T \in \mathcal{T}(S)} \left(\sum_{j=0}^{aN} \binom{N/2}{j} \cdot \left(\frac{5N}{N/2+2+j} \right)^{N/2+2+j} \cdot \frac{1}{2^j} + \sum_{j=aN+1}^{N/2} O\left(5.2852^N\right) \cdot \frac{1}{2^j} \right).$$

The terms in the first sum over j increase when $a \le 0.25$, so the sum is at most $N/2$ times its last term. Using Stirling's formula, we get that for $a \approx 0.1687$, the last term in the first sum is $\Theta^*\left(5.2852^N/2^{aN}\right) = O\left(4.7022^N\right)$. Since this also bounds the second sum, we get

$$\mathsf{st}(S) < \sum_{T \in \mathcal{T}(S)} O\left(4.7022^N\right) = O\left(4.7022^N\right) \cdot \mathsf{tr}(S),$$

as asserted. (The optimal parameter a was computed numerically.) □

Combining the bound just obtained with $\mathsf{tr}(N) < 30^N$, [26] implies

Corollary 3.3. $\mathsf{st}(N) = O\left(141.07^N\right)$.

This improves all previous upper bounds, the smallest of which is $O(158.6^N)$ [6,26].

Remark. It would be interesting to refine the bound in Theorem 2 so that it also depends on h, as in Theorem 1. An extreme situation is when S is in convex position (in which case $|F| = N-3$). In this case, it is known that $\mathsf{tr}(S) = \Theta^*(4^N)$ and $\mathsf{st}(S) = \Theta^*(6.75^N)$ (see [11]), so the exact ratio is only $\mathsf{st}(S)/\mathsf{tr}(S) = \Theta^*(1.6875^N)$. This might suggest that when h is large, the ratio should be considerably smaller, but we have not pursued this in this chapter.

Forests. For a set S of N points in the plane, we denote by $\mathcal{F}(S)$ the set of all crossing-free straight-edge forests (i.e., cycle-free graphs) of S, and put $\mathsf{f}(S) := |\mathcal{F}(S)|$. Moreover, we let $\mathsf{f}(N) = \max_{|S|=N} \mathsf{f}(S)$. Buchin and Schulz [6] have recently shown that every crossing-free straight-edge graph contains $O\left(6.4884^N\right)$ forests [improving a simple upper bound of $O^*(6.75^N)$ observed in [1]]. Following the approach of [6], we combine the bounds for spanning trees (just established) and for plane graphs with a bounded number of edges (established in Sect. 3.3), to obtain the following result.

Theorem 4. *For every set S of N points in the plane,*

$$\mathsf{f}(S) = O\left(5.3514^N\right) \cdot \mathsf{tr}(S).$$

[4]This is not quite correct: When j is close to $N/2$, the former bound is smaller [e.g., it is $O^*(5^N)$ for $j = N/2$], but we do not know how to exploit this observation to improve the bound.

Proof. We define a *k-forest* to be a forest that has k connected components. The number of k-forests of a set S is denoted by $\mathsf{f}_k(S)$. Since any spanning tree has $N-1$ edges, every k-forest has $N-k$ edges. One way to bound $\mathsf{f}_k(S)$ is by counting the number of plane graphs with $N-k$ edges. This number is bounded in Theorem 6 (from the following subsection), where the parameter c in that theorem is equal to $1-k/N$; let's denote this bound as $g_1(N,k)$. On the other hand, every k-forest is obtained by deleting $k-1$ edges from a spanning tree. This allows us to bound the number of k-forests in terms of $\mathsf{st}(S)$. Using Theorem 2, we get the bound $\mathsf{f}_k(S) \le \binom{N-1}{k-1} \cdot O\left(4.7022^N\right) \cdot \mathsf{tr}(S)$; denote this bound as $g_2(N,k)$. To bound $\mathsf{f}(S)$, we evaluate $\max_k \min\{g_1(N,k), g_2(N,k)\}$. A numerical calculation shows that the maximum value is obtained for $k' \approx 0.0285N$, and the theorem follows since $\min\{g_1(N,k'), g_2(N,k')\} = O\left(5.3514^N\right) \cdot \mathsf{tr}(S)$. □

As in the previous cases, we can combine this with the bound $\mathsf{tr}(N) < 30^N$ [26] to obtain

Corollary 3.5. $\mathsf{f}(N) = O\left(160.55^N\right)$.

Again, this should be compared with the best previous upper bound $O(194.7^N)$ [6, 26].

Consider once again the case where S consists of N points in convex position. In this case, we have $\mathsf{tr}(S) = \Theta^*(4^N)$ and $\mathsf{f}(S) = \Theta^*(8.22^N)$ (see [11]), so the exact ratio is $\mathsf{f}(S)/\mathsf{tr}(S) = \Theta^*(2.055^N)$, again suggesting that the ratio should be smaller when h is large.

3.3 The Number of Crossing-Free Straight-Edge Graphs with a Bounded Number of Edges

In this subsection, we derive upper bounds for the number of crossing-free straight-edge graphs on a set S of N points in the plane, with some constraints on the number of edges. Specifically, we bound the number of crossing-free straight-edge graphs with exactly cN edges, with at most cN edges, and with at least cN edges. The first variant has already been used in the preceding subsection for bounding the number of forests.

Crossing-Free Straight-Edge Graphs with Exactly *cN* Edges. We denote by $\mathcal{P}_c^{=}(S)$ the set of all crossing-free straight-edge graphs of S with exactly cN edges, and put $\mathsf{pg}_c^{=}(S) := |\mathcal{P}_c^{=}(S)|$. The following theorem, whose proof goes along the same lines of the proof of Theorem 1, gives a bound for $\mathsf{pg}_c^{=}(S)$.

Theorem 6. *For every set S of N points in the plane and* $0 \le c < 3$*,*

$$\mathsf{pg}_c^{=}(S) = O^*\left(B(c)^N\right) \cdot \mathsf{tr}(S),$$

where

$$B(c) := \frac{5^{5/2}}{8(c+t-1/2)^{c+t-1/2}(3-c-t)^{3-c-t}(2t)^t(1/2-t)^{1/2-t}},$$

and

$$t = \frac{1}{2}\left(\sqrt{(7/2)^2+3c+c^2}-5/2-c\right). \tag{7}$$

See Fig. 4 for a plot of the base $B(c)$ as a function of c.

Proof. The exact value of $\mathsf{pg}_c^=(S)$ is

$$\mathsf{pg}_c^=(S) = \sum_{T\in\mathcal{T}(S)} \sum_{\substack{G\in\mathcal{P}_c^=(S)\\ G\subseteq T}} \frac{1}{\mathsf{supp}(G)},$$

where $\mathsf{supp}(G)$, the support of G, is defined as in the case of general crossing-free straight-edge graphs treated in Sect. 3.1. We obtain an upper bound on this sum as follows. Consider a graph $G\in\mathcal{P}_c^=(S)$ and a triangulation $T\in\mathcal{T}(S)$, such that $G\subseteq T$. By Lemma 1.5, there is a set F of $N/2-2$ ps-flippable edges in T. Let $F_{\bar{G}}$ denote the set of edges that are in F but *not* in G, and put $j=|F_{\bar{G}}|$. As in the preceding proofs, we have $\mathsf{supp}(G)\geq 2^j$.

Next, we estimate the number of subgraphs $G\subseteq T$ for which the set $F_{\bar{G}}$ is of size j. Denote by E the set of edges of T that are not in F. As argued above, $|E|<5N/2$. To obtain a graph for which $|F_{\bar{G}}|=j$, we choose any j edges from F (the j edges of F that will not belong to G), and any subset of $cN-(N/2-2-j)=(c-1/2)N+j+2$ edges from E. If $(c-1/2)N+j+2<0$, there are no such graphs and we ignore these values of j. The number of ways to pick the edges from E is at most $O^*\left(\binom{5N/2}{(c-1/2)N+j}\right)$. This implies that

$$\begin{aligned}\mathsf{pg}_c^=(S) &< \sum_{T\in\mathcal{T}(S)}\sum_{j=0}^{N/2} O^*\left(\binom{5N/2}{(c-1/2)N+j}\cdot\binom{N/2}{j}\right)\cdot\frac{1}{2^j}\\ &= \mathsf{tr}(S)\cdot\sum_{j=0}^{N/2} O^*\left(\binom{5N/2}{(c-1/2)N+j}\cdot\binom{N/2}{j}\right)\cdot\frac{1}{2^j}.\end{aligned} \tag{8}$$

[As already noted, when $c<1/2$, only the terms for which $(c-1/2)N+j\geq 0$ are taken into account.]

As in the preceding subsection, it suffices to consider only the largest term of the sum. For this, we consider the quotient of the jth and $(j-1)$st terms [ignoring the $O^*(\cdot)$ notation, which will not affect the exponential order of growth of the terms], which is

$$\frac{\binom{N/2}{j}\binom{5N/2}{(c-1/2)N+j}}{2\binom{N/2}{j-1}\binom{5N/2}{(c-1/2)N+j-1}} = \frac{\left(N/2-j+1\right)\left(5N/2-(c-1/2)N-j+1\right)}{2j\left((c-1/2)N+j\right)}.$$

To simplify matters, we put $a = N/2$ and $b = (c - 1/2)N$. Moreover, since we are only looking for an asymptotic bound, and are willing to incur small multiplicative errors within the $O^*(\cdot)$ notation, we may ignore the two $+1$ terms in the numerator when N is sufficiently large; we omit the routine algebraic justification of this statement. The above quotient then becomes (approximately) $\dfrac{(a-j)(5a-b-j)}{2j(b+j)}$, which is larger than 1 whenever

$$\begin{aligned} j &< \tfrac{1}{2}\left(\sqrt{56a^2 + 8ab + b^2} - 6a - b\right) \\ &= \tfrac{N}{2}\left(\sqrt{(7/2)^2 + 3c + c^2} - 5/2 - c\right) = tN, \end{aligned}$$

with t given in (7). A simple calculation shows that $0 \le t < 1/2$ and $0 \le c - 1/2 + t \le 5/2$ for $0 \le c \le 3$. In other words (and rather unsurprisingly), the index $j = tN$ attaining the maximum does indeed lie in the range where the two binomial coefficients in the corresponding terms in (8) are both well defined (nonzero).

Now that we have the largest term of the sum in (8), we obtain

$$\mathsf{pg}_c^{=}(S) = \mathsf{tr}(S) \cdot O^*\left(\binom{5N/2}{(c-1/2)N + tN} \cdot \binom{N/2}{tN} \cdot \frac{1}{2^{tN}}\right).$$

Using Stirling's approximation, we have

$$\begin{aligned} \mathsf{pg}_c^{=}(S) &= \mathsf{tr}(S) \cdot O^*\left(\left(\frac{(5/2)^{5/2}}{(c+t-1/2)^{c+t-1/2}(3-c-t)^{3-c-t}} \cdot \frac{(1/2)^{1/2}}{t^t(1/2-t)^{1/2-t}} \cdot \frac{1}{2^t}\right)^N\right) \\ &= \mathsf{tr}(S) \cdot O^*\left(\left(\frac{5^{5/2}}{8(c+t-1/2)^{c+t-1/2}(3-c-t)^{3-c-t}(2t)^t(1/2-t)^{1/2-t}}\right)^N\right), \end{aligned}$$

as asserted. □

Crossing-Free Straight-Edge Graphs with at Most *cN* Edges. For a set S of N points in the plane and a constant $0 < c < 3$, we denote by $\mathcal{P}_c^{\le}(S)$ the set of all crossing-free straight-edge graphs of S with at most cN edges, and put $\mathsf{pg}_c^{\le}(S) := \left|\mathcal{P}_c^{\le}(S)\right|$. The bound for $\mathsf{pg}_c^{=}(S)$ in Theorem 6 helps us to determine the bound for $\mathsf{pg}_c^{\le}(S)$.

Theorem 7. *For every set S of N points in the plane and $0 < c < 3$,*

$$\mathsf{pg}_c^{\le}(S) = \begin{cases} O^*\left(B(c)^N\right) \cdot \mathsf{tr}(S) & \text{if } c \le 19/12, \\ O^*\left((4\sqrt{3})^N\right) \cdot \mathsf{tr}(S) & \text{otherwise}, \end{cases}$$

where $B(c)$ is defined as in Theorem 6.

Proof. We begin by noticing that

$$\mathsf{pg}_c^{\le}(S) = \sum_{0<j\le cN} \mathsf{pg}_{j/N}^{=}(S) = O^*\left(\max_{c'\le c} \mathsf{pg}_{c'}^{=}(S)\right) = O^*\left(\max_{c'\le c} B(c')^N\right)\cdot \mathsf{tr}(S). \quad (9)$$

Let $F(c)$ be the natural logarithm of the nonconstant part of the denominator of $B(c)$ (the numerator is a constant). That is,

$$F(c) = (c+t-1/2)\ln(c+t-1/2) + (3-c-t)\ln(3-c-t) + t\ln(2t) + (1/2-t)\ln(1/2-t).$$

Since each of the terms in the logarithms is positive when $0 < c < 3$, finding a maximum for $B(c)$ in this range is equivalent to finding a minimum for $F(c)$. Put $t' = t'(c)$ (the derivative of t as a function of c). Then

$$\begin{aligned}F'(c) &= (1+t')\ln(c+t-1/2) + (1+t') - (1+t')\ln(3-c-t) - (1+t')\\ &\quad + t'\ln(2t) + t' - t'\ln(1/2-t) - t'\\ &= (1+t')\ln\left(\frac{c+t-1/2}{3-c-t}\right) + t'\ln\left(\frac{2t}{1/2-t}\right).\end{aligned}$$

For $c = 19/12$ we have $t = 1/6$ and the arguments of both logarithms are 1, so $F'(c) = 0$. Easy calculations show that

$$t' = \frac{1}{2}\left(\frac{3/2+c}{\sqrt{(7/2)^2+3c+c^2}} - 1\right) = -\frac{0.5+t}{\sqrt{(7/2)^2+3c+c^2}}.$$

This is easily seen to imply that $t' \le 0$ and $1+t' \ge 0$ for $0 < c < 3$. This implies that $c+t$ is monotone increasing with c, and that t is decreasing (because $1+t' \ge 0$ and $t' \le 0$). It follows that $F'(c)$ is increasing with c, implying that $F(c)$ attains its minimum at $c = 19/12$ [since $F'(19/12) = 0$]. Another easy calculation shows that $B(19/12) = 4\sqrt{3}$. □

For example, Theorem 7 implies that there are at most $O^*\left(5.4830^N\right)\cdot \mathsf{tr}(S)$ crossing-free straight-edge graphs with at most N edges, over any set S of N points in the plane. In particular, this is also an upper bound on the number of crossing-free straight-edge forests on S, or of spanning trees, or of spanning cycles. Of course, better bounds exist for these three special cases, as demonstrated earlier in this chapter for the first two bounds.

Crossing-Free Straight-Edge Graphs with at Least cN Edges. We next bound the number of plane graphs with at least cN edges. Following the notations used above, we denote by $\mathcal{P}_c^{\ge}(S)$ the set of all crossing-free straight-edge graphs of S with at least cN edges, and put $\mathsf{pg}_c^{\ge}(S) := \left|\mathcal{P}_c^{\ge}(S)\right|$.

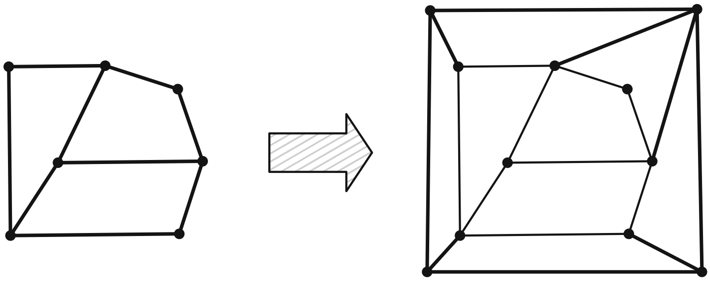


Fig. 9 A quadrangulation of S and a quadrangulation of S' that contains it

Theorem 8. *For every set S of N points in the plane and* $0 < c < 3$*,*

$$\mathsf{pg}^{\geq}_{c}(S) = \begin{cases} O^*\left(B(c)^N\right) \cdot \mathsf{tr}(S) & \text{if } c \geq 19/12, \\ O^*\left((4\sqrt{3})^N\right) \cdot \mathsf{tr}(S) & \text{otherwise}. \end{cases}$$

Proof. Similar to (9), we can bound $\mathsf{pg}^{\geq}_{c}(S)$ by

$$\mathsf{pg}^{\geq}_{c}(S) = \sum_{cN \leq j < 3N} \mathsf{pg}^{=}_{j/N}(S) = O^*\left(\max_{c \leq c' < 3} \mathsf{pg}^{=}_{c'}(S)\right) = O^*\left(\max_{c \leq c' < 3} B(c)^N\right) \cdot \mathsf{tr}(S). \tag{10}$$

The analysis of (10) is symmetric to the one presented in the proof of Theorem 7, and the theorem follows. □

As an application, consider the problem of bounding the number of quadrangulations of S, namely, crossing-free straight-edge connected graphs on the vertex set S with no isolated vertices, that include all the hull edges of $\mathrm{conv}(S)$, and where every bounded face is a quadrilateral. When h is odd, no quadrangulations can be embedded over S (e.g., see [4]). We may thus assume that h is even, and create a superset $S' \supset S$ as follows. We take a quadrilateral Q that contains S in its interior, and add the vertices of Q to S. It is easy to see that every quadrangulation of S is contained in at least one quadrangulation of S'; see Fig. 9 for an illustration.[5] Therefore, it suffices to bound the number of quadrangulations of S'.

Using Euler's formula, we notice that a quadrangulation of S' has $N+1$ quadrilaterals, and $2N+4$ edges (since $|S'| = N+4$). Therefore, we can use Theorem 7 with $c = 2$, which implies a bound of $O^*(6.1406^N) \cdot \mathsf{tr}(N) = O(184.22^N)$. [There are actually $N+4$ points and $c = (2N+4)/(N+4)$. However, since we are

[5]We need to construct a quadrangulation of the annulus-like region between Q and the convex hull of S. We start by connecting a vertex of Q to a vertex of the convex hull, and in each step we add a quadrangle by either marching along two edges of the hull or along one edge of the hull and one edge of Q. This produces the desired quadrangulation.

only interested in the exponential part of the bound, the above bound, with the $O^*(\cdot)$ notation, does hold.] We are not aware of any previous explicit treatment of this problem.

4 Conclusion

In this chapter, we have introduced the notion of pseudo-simultaneously flippable edges in triangulations, have shown that many such edges always exist, and have used them to obtain several refined bounds on the number of crossing-free straight-edge graphs on a fixed (labeled) set of N points in the plane. The chapter raises several open problems and directions for future research.

One such question is whether it is possible to further extend the notion of ps-flippability. For example, one could consider, within a fixed triangulation T, the set of diagonals of a collection of pairwise interior-disjoint simple, but not necessarily convex, polygons. The number of such diagonals is likely to be larger than the size of the maximal set of ps-flippable edges, but it not clear how large the number of triangulations is that can be obtained by redrawing diagonals.

We are currently working on two extensions to this work. The first extends our techniques to the cases of crossing-free straight-edge perfect matchings and spanning (Hamiltonian) cycles. This is done within the linear algebra framework introduced by Kasteleyn (see [5, 20]) and can be found in [28]. The second work studies charging schemes in which the charge is moved across certain objects belonging to different crossing-free straight-edge graphs over the same point set. This cross-graph charging scheme allows us to obtain bounds that do not depend on the current upper bound for $\mathtt{tr}(N)$ [and bounds that depend on $\mathtt{tr}(N)$ in a nonlinear fashion].

Acknowledgements Work on this chapter by Micha Sharir and Adam Sheffer was partially supported by Grant 338/09 from the Israel Science Fund. Work by Micha Sharir was also supported by NSF Grant CCF-08-30272, by Grant 2006/194 from the U.S.-Israel Binational Science Foundation, and by the Hermann Minkowski-MINERVA Center for Geometry at Tel Aviv University. Work by Csaba D. Tóth was supported in part by NSERC Grant RGPIN 35586. Research by this author was conducted at ETH Zürich. Emo Welzl acknowledges support from the EuroCores/EuroGiga/ComPoSe SNF Grant 20GG21_134318/1. Part of the work on this chapter was done at the Centre Interfacultaire Bernoulli (CIB), EPFL, Lausanne, during the Special Semester on Discrete and Computational Geometry, Fall 2010, and was supported by the Swiss National Science Foundation.

References

1. O. Aichholzer, T. Hackl, C. Huemer, F. Hurtado, H. Krasser, B. Vogtenhuber, On the number of plane geometric graphs. Graphs Comb. **23**(1), 67–84 (2007)
2. M. Ajtai, V. Chvátal, M.M. Newborn, E. Szemerédi, Crossing-free subgraphs. Ann. Discr. Math. **12**, 9–12 (1982)

3. P. Bose, F. Hurtado, Flips in planar graphs. Comput. Geom. Theor. Appl. **42**(1), 60–80 (2009)
4. P. Bose, G. Toussaint, No quadrangulation is extremely odd, in *Algorithms and Computations*. Lecture Notes in Computer Science, vol. 1004 (Springer-Verlag, Berlin, 1995), pp. 372–381
5. K. Buchin, C. Knauer, K. Kriegel, A. Schulz, R. Seidel, On the number of cycles in planar graphs, in *Proceedings of the 17th Computing and Combinatorics Conference*. Lecture Notes Computer Science, vol. 4598 (Springer, Berlin, 2007), pp. 97–107
6. K. Buchin, A. Schulz, On the number of spanning trees a planar graph can have, in *Proceedings of the 18th Annual European Symposium on Algorithms*. Lecture Notes Computer Science, vol. 6346 (Springer, Berlin, 2010), pp. 110–121
7. J.A. De Loera, J. Rambau, F. Santos, *Triangulations: Structures for Algorithms and Applications* (Springer, Berlin, 2010)
8. M.O. Denny, C.A. Sohler, Encoding a triangulation as a permutation of its point set, in *Proceedings of the 9th Canadian Conference on Computational Geometry*, 1997, Kingston, Ontario, Canada, pp. 39–43
9. A. Dumitrescu, A. Schulz, A. Sheffer, Cs.D. Tóth, Bounds on the maximum multiplicity of some common geometric graphs, in *Proceedings of the 28th Symposium on Theoretical Aspects of Computer Science*, 2011, vol. 5 of LIPICS, Schloss Dagstuhl, Dagstuhl, pp. 637–648
10. I. Fáry, On straight line representations of planar graphs. Acta Sci. Math. (Szeged) **11**, 229–233 (1948)
11. P. Flajolet, M. Noy, Analytic combinatorics of non-crossing configurations. Discr. Math. **204**, 203–229 (1999)
12. S. Fortune, Voronoi diagrams and Delaunay triangulations, in *Computing in Euclidean Geometry*, ed. by D.A. Du, F.K. Hwang (World Scientific, New York, 1992), pp. 193–233
13. J. Galtier, F. Hurtado, M. Noy, S. Pérennes, J. Urrutia, Simultaneous edge flipping in triangulations. Int. J. Comput. Geom. Appl. **13**(2), 113–133 (2003)
14. A. García, M. Noy, J. Tejel, Lower bounds on the number of crossing-free subgraphs of K_N. Comput. Geom. Theor. Appl. **16**(4), 211–221 (2000)
15. J. García-Lopez, M. Nicolás, Planar point sets with large minimum convex partitions. *Abstracts 22nd European Workshop on Computational Geometry*, 2006, Delphi, Greece, pp. 51–54
16. O. Giménez, M. Noy, Asymptotic enumeration and limit laws of planar graphs. J. Am. Math. Soc. **22**, 309–329 (2009)
17. Ø. Hjelle, M. Dæhlen, *Triangulations and Applications* (Springer, Berlin, 2009)
18. K. Hosono, On convex decompositions of a planar point set. Discr. Math. **309**, 1714–1717 (2009)
19. F. Hurtado, M. Noy, J. Urrutia, Flipping edges in triangulations. Discr. Comput. Geom. **22**, 333–346 (1999)
20. L. Lovász, M. Plummer, Matching Theory (North Holland, Budapest, 1986)
21. R.C. Mullin, On counting rooted triangular maps. Can. J. Math. **7**, 373–382 (1965)
22. A. Razen, J. Snoeyink, E. Welzl, Number of crossing-free geometric graphs vs. triangulations. Electron. Notes Discr. Math. **31**, 195–200 (2008)
23. A. Ribó Mor, Realizations and counting problems for planar structures: trees and linkages, polytopes and polyominos. Ph.D. thesis, Freie Universität, Berlin, 2005
24. G. Rote, The number of spanning trees in a planar graph. Oberwolfach Rep. **2**, 969–973 (2005)
25. F. Santos, R. Seidel, A better upper bound on the number of triangulations of a planar point set. J. Comb. Theor. Ser. A **102**(1), 186–193 (2003)
26. M. Sharir, A. Sheffer, Counting triangulations of planar point sets. Electron. J. Comb. **18**(1), P70 (2011)
27. M. Sharir, A. Sheffer, E. Welzl, On degrees in random triangulations of point sets. J. Comb. Theor. A **118**, 1979–1999 (2011)
28. M. Sharir, A. Sheffer, E. Welzl, Counting plane graphs: perfect matchings, spanning cycles, and Kasteleyn's technique. *Proc. 28th ACM Symp*. on Computational Geometry, 2012, ACM, New York, pp. 189–198
29. M. Sharir, E. Welzl, Random triangulations of planar point sets, in *Proceedings of the 22nd ACM Symposium on Computational Geometry*, 2006, ACM, New York, pp. 273–281

30. D.L. Souvaine, C.D. Tóth, A. Winslow, Simultaneously flippable edges in triangulations, in *Proceedings of the XIV Spanish Meeting on Computational Geometry*, 2011, Alcalá de Henares, Spain, pp. 137–140
31. R.P. Stanley, *Enumerative Combinatorics*, vol. 2 (Cambridge University Press, Cambridge, 1999)
32. W.T. Tutte, A census of planar maps. Can. J. Math. **15**, 249–271 (1963)
33. J. Urrutia, Open problem session, in *Proceedings of the 10th Canadian Conference on Computational Geometry*, McGill University, Montréal, 1998
34. K. Wagner, Bemerkungen zum Vierfarbenproblem. J. Deutsch. Math.-Verein. **46**, 26–32 (1936)

Plane Geometric Graph Augmentation: A Generic Perspective

Ferran Hurtado and Csaba D. Tóth

Abstract Graph augmentation problems are motivated by network design and have been studied extensively in optimization. We consider augmentation problems over plane geometric graphs, that is, graphs given with a crossing-free straight-line embedding in the plane. The geometric constraints on the possible new edges render some of the simplest augmentation problems intractable, and in many cases only extremal results are known. We survey recent results, highlight common trends, and gather numerous conjectures and open problems.

1 Introduction

Let $G = (V, E)$ be a graph. We say that a second graph $G' = (V, E \cup E')$ obtained by adding a set E' of edges to G is an *(edge) augmentation* of G. The goal of this operation is to ensure that the augmented graph G' has some desired property. Usually, one would like to achieve the goal at a minimum cost, which is typically measured by the *number* of new edges, although weighted versions are also possible. In this survey, we consider edge augmentation only, but we note that in general one could augment a graph with both new *vertices* and edges, or even *subdivide* an edge (by replacing an edge with a path).

A *geometric graph* $G = (V, E)$ is a graph drawn in the plane such that the vertex set V is a set of points in the plane, and the set of edges E consists of line segments with endpoints in V, whose relative interiors are disjoint. Two edges of a

F. Hurtado
Departament de Matemàtica Aplicada II, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
e-mail: ferran.hurtado@upc.edu

C.D. Tóth (✉)
Department of Mathematics and Statistics, University of Calgary, Calgary, AB, Canada
e-mail: cdtoth@ucalgary.ca

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_17,

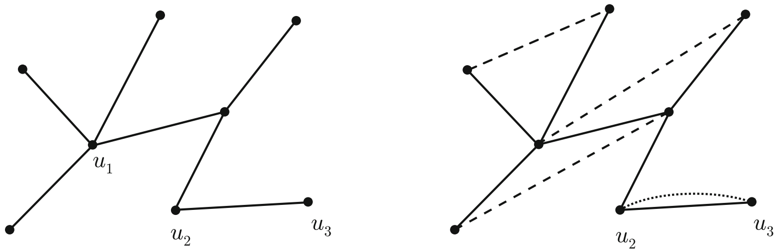


Fig. 1 The tree on the left does not contain any perfect matching. Augmenting the tree with the three *dashed* edges on the right results in a graph that contains a perfect matching. For example, the three new edges together with the original edge u_2u_3 form a perfect matching

geometric graph *cross* if they have an intersection point lying in the relative interior of both edges. We consider *crossing-free* (or *noncrossing*) geometric graphs, where no two edges cross. The terms *plane geometric graph*, *(crossing-free) segment configuration*, and *planar straight-line graph* (for short PSLG) will also be used here as synonyms for crossing-free geometric graphs. Rather than using only one of these terms in this panoramic chapter, we use them all interchangeably, as otherwise a reader who follows the references may be confused by the diversity of the terminology in the literature.

Compatibility and Visibility. Two crossing-free geometric graphs, $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, are *compatible* if their union $(V_1 \cup V_2, E_1 \cup E_2)$ is also a crossing-free geometric graph. In this survey we focus on augmentation problems in which a geometric graph $G = (V, E)$ is augmented to a graph compatible with G. For example, G could be a set of disjoint segments and we may want to add new segments among the endpoints in order to obtain a crossing-free spanning tree with certain desirable properties. Alternatively, we may be given an arbitrary noncrossing geometric graph G and we might want to add the minimum number of edges to increase its vertex or edge connectivity. In a third example, we may consider whether from any given plane spanning tree G we can construct an augmentation G' containing a Hamiltonian cycle or, when V is even, a perfect matching (Fig. 1).

The possible new edges that may be added to a geometric graph $G = (V, E)$ can be interpreted in the context of *visibility* problems, which emerged in the late 1980s and early 1990s. We say that two vertices $p, q \in V$ *see* each other (i.e., they are mutually *visible*) if the segment pq does not cross any edge in E and its relative interior does not contain any vertex in V. The edges in E together with all visibility edges form a geometric graph, called the *segment endpoint visibility graph*, for the segment configuration E. Notice that a segment between two vertices in V belongs to this graph if and only if it is compatible with E. We refer the readers to the surveys [11, 47, 80, 107] for properties and related results stated in the visibility framework.

For a set of points S in general position (that is, no three points on a line), we denote by $K(S)$ the complete geometric graph on vertex set S, that is, where every pair of points is joined by an edge.

A Unifying Framework. The concepts of augmentation and compatibility allow us to describe, under a unifying framework, several problems that have a common flavor and have attracted substantial research, yet make use of different terminology or aim at different goals. A unified view helps to identify and understand common methods and common difficulties.

Our survey is not intended to be exhaustive, since there are many different variants of the augmentation problems under this framework. Rather, we have selected a sample of representative problems that provide the reader with a global comprehension of the field. Along the way, we also present several open problems and unsettled conjectures.

In the augmentation problems we consider, we are typically given a planar straight-line graph G and a property $\mathcal{P}$; and our goal is to find an augmentation G' with property $\mathcal{P}$. When the augmentation is not necessarily feasible, one may seek efficient decision algorithms and combinatorial characterizations. If the augmentation is feasible, then the goal is to find the minimum number of new edges required. This includes possible combinatorial characterizations of the minimum number, efficient algorithms to construct a minimum augmentation, approximation algorithms, as well as bounds on the extremal number of edges required over certain classes of graphs.

2 An Introductory Example: Augmenting a Matching

A set of disjoint line segments in the plane is, in effect, a crossing-free straight-line drawing of a perfect matching, where the segment endpoints are the vertices. This is a basic scenario for augmentation problems, in which we can neatly see geometry and graph theory interplay. In this section, we review some fundamental results about this particular case.

2.1 From Perfect Matchings to Hamiltonian Cycles and Paths

A *polygonization* of a given point set S is a simple polygon whose vertex set is exactly S, in other words, a crossing-free spanning cycle in $K(S)$. It is easy to see that every set of $n \geq 3$ points in general position admits a polygonization (e.g., every minimum Euclidean TSP is crossing-free). Consider now a set M of n disjoint line segments, let us denote by S_M the set of their endpoints, and assume that S_M is in general position. If the matching M can be augmented with $|M|$ edges to a noncrossing cycle P (i.e., a simple polygon), then every other edge of P belongs to M. Such a polygon P is also called an *alternating polygonization* of M. It

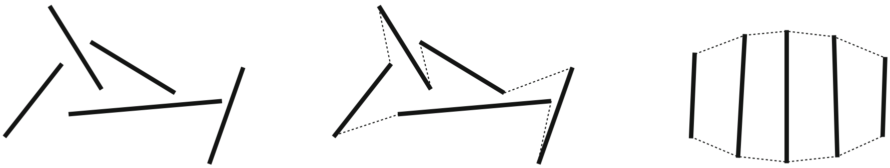


Fig. 2 A set of five segments (*left*), which admits an alternating polygonization (*middle*). This is not possible for the five segments on the *right* of the figure because their endpoints are in convex position, and the only simple polygonization they admit is the boundary of their convex hull, missing the three central segments

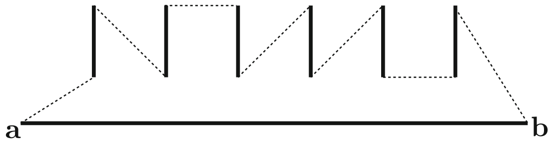


Fig. 3 For this set of $n=7$ segments we can easily construct 2^{n-1} different alternating polygonizations. Imagine that we travel from a to b, traversing the $n-1$ vertical segments in our trajectory. At every departure we can choose whether we arrive at the top or at the bottom endpoint of the next vertical segment

is easy to see that M does not always admit an alternating polygonization (see Fig. 2). Rappaport [84] proved that it is NP-complete to decide whether a plane geometric graph can be augmented to a spanning cycle; and it is also conjectured to be NP-complete to decide whether a noncrossing matching admits an alternating polygonization. Studying this problem had been suggested by Toussaint around 1985. Rappaport et al. [85] proved that the decision problem is polynomially solvable in some special cases, for example, when the segments in M are *convexly independent*; that is, each of them has at least one endpoint in the convex hull $\mathrm{CH}(S_M)$.

Let us remark that the noncrossing geometric constraint is the core of the difficulty of this problem. If we disregard crossings, we are simply in the combinatorial scenario, and any perfect matching M in the complete graph K_{2n} can be augmented in $2^n(n-1)!$ ways to a Hamiltonian cycle in which every other edge belongs to M. For embedded geometric graphs, we have seen an example of segments not admitting any alternating polygonization (Fig. 2). When they exist, it is possible that there is only one, for example, if M consists of every other edge of a convex $2n$-gon. But there can be exponentially many alternating polygonizations, as in the configuration in Fig. 3. Nevertheless, their number is bounded above by c^n for some constant c (see Sect. 3), in contrast to more than $n!$ different alternating cycles among abstract graphs.

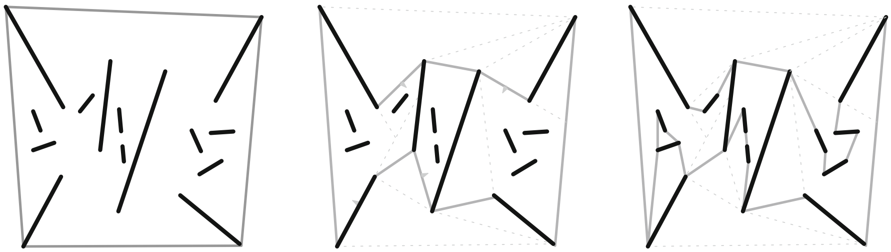


Fig. 4 A set M of disjoint line segments, where P is initially the boundary of the convex hull $\mathrm{CH}(S_M)$ (*left*). Polygon P is extended until it contains either both endpoints or neither endpoint of each segment. Every segment in $M \setminus P$ lies in a convex tile adjacent to a unique edge of P, as indicated by small arrow heads (*middle*). In each convex tile containing noncollinear segments, there is a compatible spanning cycle by induction, which can be fused to P via the corresponding edge of P (*right*)

We have seen that a noncrossing matching does not necessarily admit a Hamiltonian cycle. Mirzaian [74] conjectured that every noncrossing matching with noncollinear vertices can be augmented to a Hamiltonian plane geometric graph (that is, the augmented graph contains a Hamiltonian cycle, but this cycle does not have to contain M). Several years later, Hoffmann and Tóth [56] confirmed this conjecture in the affirmative. They constructed a noncrossing cycle P incrementally, starting from the convex hull $\mathrm{CH}(S_M)$. The polygon P is then successively extended to pass through more segment endpoints, while it remains compatible with M. In a first phase of their algorithm, P is incrementally extended to include the second endpoint of every segment that already has an endpoint along P (similar to Mirzaian's technique to handle convexly independent segments [74]); and simultaneously they constructed a tiling of $\mathrm{CH}(S_M)$ such that every nonempty tile is adjacent to a unique edge of P. At the end of the first phase, all tiles are convex, and every remaining edge in $M \setminus P$ lies in the interior of some tile. In the second phase, a Hamiltonian polygon is computed, by induction, in each tile that contains noncollinear segments of M. The small polygons in the tiles, as well as collinear segments in some other tiles, are fused to P by modifying the common edge of the tile with P (see Fig. 4 for an example).

Mirzaian [74] also made a slightly stronger conjecture: that every noncrossing matching M with noncollinear vertices would admit a *circumscribing polygon*, that is, a polygonization of S_M in which every segment of M would be either an edge or an internal diagonal of the polygon (external diagonals are excluded). He proved this property for convexly independent segments. However, Urabe and Watanabe found a counterexample to this conjecture [106]. O'Rourke and Rippel found a new family of segments admitting a circumscribing polygonization, namely, sets of unit segments such that the line containing each segment misses all the other segments [81].

Pach and Rivera proved that every set M of n segments has a subset $M' \subseteq M$ of size roughly $\sqrt[3]{n}$, such that M' admits a circumscribing polygon [82]. Let us note,

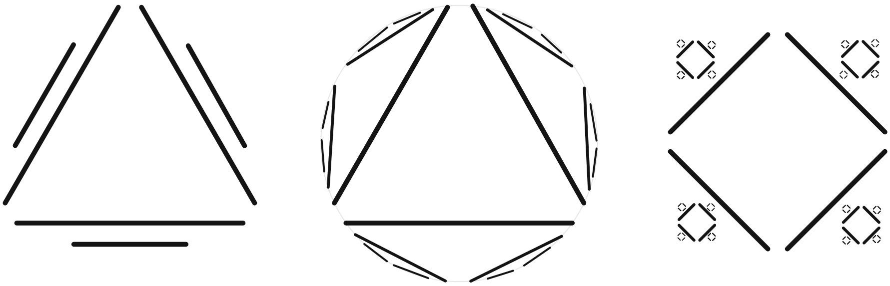


Fig. 5 The set of segments that does not admit a spanning alternating path (*left*). This idea can be extended to a tree-like construction (*middle*) in which the endpoints of the segments are in convex position, in such a way that every segment "hides" two subtrees having the same structure. Only a subset of logarithmic size can be included in any alternating simple path. A similar construction (*right*) with orthogonal segments arranged in a 4-ary tree

though, that this polygon may not be compatible with M, it may cross some edges in $M \setminus M'$. If we instead insist on alternating polygons that are compatible with all segments in M, then the best result one can prove is that there are always two segments in M that can be augmented to a simple quadrilateral compatible with M (M.E. Houle, private communication). The bound of at most 2 is tight, as shown by the example in Fig. 2, right.

Around 1992, Urrutia asked what is the maximum length of a compatible alternating path for a perfect matching M? It is a significant relaxation to require an *open* polygonal path rather than a closed polygon. For example, the segments on the right of Fig. 2, in fact, can be augmented to an alternating Hamiltonian path. However, such a path is not always possible (Fig. 5, left), and it is not difficult to construct sets of n segments in which any noncrossing alternating path compatible with M has $O(\log n)$ length (Fig. 5, middle). Hoffmann and Tóth [55] proved that for every set of n disjoint segments, there is an alternating path of length $\Omega(\log n)$ compatible with M, hence showing that the upper bound is tight up to a multiplicative constant. For n disjoint orthogonal line segments, Tóth [103] proved an asymptotically tight bound of $\Theta(\sqrt{n})$ (Fig. 5, right).

2.2 From Perfect Matchings to Encompassing Trees

We saw in the previous section that not every perfect matching M can be augmented to a Hamiltonian cycle. It is not very difficult to prove, though, that there is always an *encompassing tree*, that is, a compatible spanning tree, that contains M as a subgraph: One can construct an encompassing tree from any triangulation of M (Fig. 6). Nevertheless, a naïve algorithm might yield some high-degree node in the tree. Is it always possible to find an encompassing tree of bounded degree?

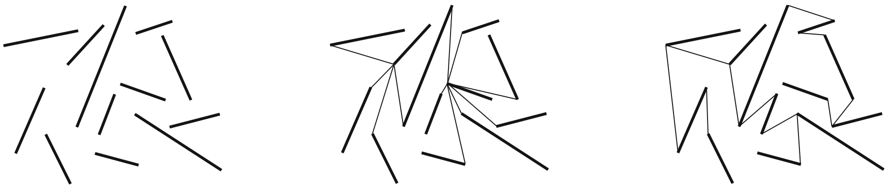


Fig. 6 The set of 12 segments (*left*), which can be encompassed by a tree in which the maximum degree of a node is 8 (*middle*), but also by a binary tree (*right*)

Bose and Toussaint [20] proved in 1992 that every n disjoint segments admit an encompassing tree with maximum degree at most 7, and gave an $O(n \log n)$ algorithm for its construction. Shortly afterward, they improved on this result, showing that every set of disjoint segments can be encompassed by a spanning tree of maximum degree 3 (a binary tree), which can be constructed in optimal $\Theta(n \log n)$ time [18]. The bound on the degree is the best possible, because an encompassing tree with maximum degree 2 would be an alternating Hamiltonian path, which we know is not always possible. Later Souvaine and Tóth [98] obtained a generalization: Every (disconnected) PSLG on n vertices can be augmented into a connected PSLG such that the degree of each vertex increases by at most 2, and the augmentation can be computed in $O(n \log n)$ time.

A colored version of this problem was posed in 2004 in the conference version of [62]: Given a set of disjoint segments with a proper vertex coloring (two endpoints of each segments are colored differently), is it always possible to find a plane-encompassing tree such that the new edges also respect the coloring? Hurtado et al. [62] gave an affirmative answer. Hoffmann and Tóth [54] later proved that there is always a *binary* encompassing tree respecting the initial coloring of the segment endpoints. This result was also generalized for arbitrary segment configurations [57]: Every vertex-colored PSLG without singleton components can be augmented into a connected PSLG such that the degree of each vertex increases by at most two and the new edges respect the coloring. The exclusion of singleton components is necessary, since it is possible that a singleton is visible only from vertices of the same color (Fig. 7, left).

Another variation arose in the context of *pseudo-triangulations*, a topic that spurred substantial attention in the first decade of the 21st century (see the survey [87]), especially after the works [99] and [100] by Streinu. A *pseudo-triangle* is a simple polygon having exactly three convex vertices, and a *pseudo-triangulation* of a set S of n points is a noncrossing geometric graph, whose bounded faces decompose the convex hull $\mathrm{CH}(S)$ into pseudo-triangle faces. While the number of triangles and edges in a triangulation of S depends on the number of vertices on $\mathrm{CH}(S)$, it is always possible to decompose $\mathrm{CH}(S)$ into exactly $n-2$ pseudo-triangles, using exactly $2n-3$ edges, if S is in general position, no matter how many points of S lie in the interior of the convex hull. These decompositions minimize the number of pseudo-triangles (notice that a triangulation is also a

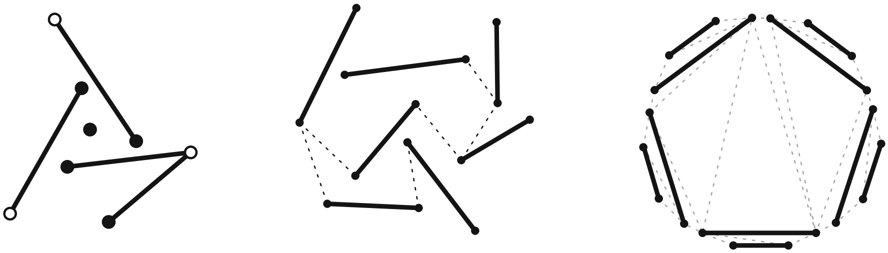


Fig. 7 A vertex-colored graph where a singleton vertex is only visible from vertices of the same color (*left*). Disjoint line segments can be augmented to a pointed binary spanning tree (*middle*). A set of 10 segments, where for every pseudo-triangulation there is a vertex of degree at least 7 (*right*)

pseudo-triangulation) and have the property that every vertex $p \in S$ is *pointed*: One of the faces incident to p has a reflex angle at p. The pointedness property is related to the rigidity of graphs [100], which is one of the reasons for the interest in pseudo-triangulations. Motivated by this framework, Hoffmann et al. [53] proved that every set of disjoint segments has an encompassing tree in which every vertex is pointed and has degree at most 3 (Fig. 7, middle). Such a pointed binary tree can be augmented to a minimal pseudo-triangulation, and building on that they also proved that every set of disjoint line segments in the plane has an encompassing minimal pseudo-triangulation whose maximum vertex degree is bounded by 7, and this bound cannot be improved (Fig. 7, right).

2.3 Looking for a Second Matching

A variant close in spirit to the original problem on alternating polygonizations of a matching was posed by Aichholzer et al. [2]. Every alternating polygonization has an even number of edges and is the union of two disjoint perfect matchings, M and M'. On the other hand, if M and M' are two disjoint and compatible perfect matchings of a point set, then their union is a set of simple polygons [cycles in $K(S_M)$], each of which has an even number of edges. Let us observe that the matching in the right part of Fig. 2 has an odd number of edges and does not admit any compatible disjoint perfect matching. Several other constructions were given in [2] for an odd number of disjoint segments without any compatible disjoint perfect matchings. The authors conjectured that for every noncrossing perfect matching M with an even number of edges in general position, there is a disjoint compatible perfect matching (*disjoint compatible matching conjecture*).

The positive results in [2] toward a solution were only partial. They confirmed the conjecture in some special cases (namely, for convexly independent or orthogonal segments), and they also proved that there is always a set of alternating polygons, all of them compatible with M, encompassing at least $4/5$ of the segments in M.

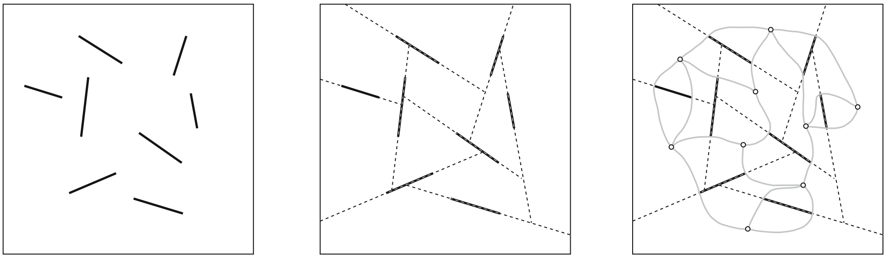


Fig. 8 The segments on the *left* part can be extended to produce a convex subdivision of the plane, as shown in the *middle*. The dual multigraph of this subdivision is shown on the *right*

On the other hand, the techniques used for these results led the authors to pose stronger conjectures, involving the *convex subdivision* of the free space around the line segments. In fact, this geometric tool has been used in solving nearly all the problems described in this section. One easy way to construct a convex subdivision for the segments in M is the following (refer to Fig. 8). For each endpoint q of a segment $pq = s \in M$, extend s along the ray $\overrightarrow{pq}$ beyond q until it hits another segment, a previous extension, or (if it is not blocked) to infinity. Different orders in which the extensions are drawn may yield different subdivisions, but in all cases the plane is subdivided into $n+1$ convex cells, where $n = |M|$. Clearly, not all convex subdivisions can be obtained in this way; e.g., the minimum number of cells in a convex subdivision may be far fewer than $n+1$. In the *dual multigraph* D associated with a convex subdivision, the vertex set $V(D)$ is the set of convex cells, and every segment endpoint p corresponds to an edge in $E(D)$ between two cells incident to p (double edges are possible if a segment in M lies on the common boundary between two cells).

The additional conjectures in [2] were stated in the terms of the dual multigraph D associated with a "suitable" convex subdivision. Orienting an edge of D toward a node $v \in V(D)$ can be seen as assigning a segment endpoint to the cell corresponding to v. In an *even orientation* of an undirected (multi-)graph, edges are oriented in such a way that every node has an even in-degree. In particular, an even orientation of D means that an even number of segment endpoints are assigned to each convex cell of the subdivision. Now, it is not difficult to see that in this case, one can match the endpoints assigned to each cell using disjoint segments lying in the interior of the cell (and hence compatible with M), with the only exceptional case that a cell is assigned to exactly two segment endpoints, which are the two endpoints of the same segment of M. We can encode this information in the dual graph: Two adjacent edges of D are said to be in *conflict* if they correspond to the two endpoints of a segment in M; and an even orientation of D is *conflict-free* if no two conflicting edges are oriented into a node of in-degree 2. Hence, if M admits a convex subdivision where the dual graph has a conflict-free even orientation, this immediately implies the existence of a disjoint compatible matching M'.

A solution to the disjoint compatible matching conjecture has been claimed very recently by Ishaque et al. [64] and follows the approach described in the preceding paragraph. For every set of $n \geq 2$ disjoint line segments in the plane in general position, they construct a convex subdivision such that the associated dual graph D contains two edge-disjoint spanning trees. The subdivision is obtained by an iterative process, in which the number of cells may drop below $n+1$, but the number of edges remains $2n$, since they are in bijection with the segment endpoints. They also show that every multigraph that has an even number of edges and contains two edge-disjoint spanning trees must have a conflict-free even orientation. Based on the argument in the previous paragraph, this already implies the disjoint compatible matching conjecture.

None of the stronger conjectures formulated by Aichholzer et al. [2] has been confirmed yet, but some of them have been refuted. For example, if we insist that the convex subdivision must be constructed by extending the segments successively beyond their endpoints (as in Fig. 8), then the dual graph does not always contain two edge-disjoint spanning trees [7]. In the proof of Ishaque et al. [64], it was essential to work with a broader class of convex subdivisions, which may have fewer than $n+1$ convex cells.

3 An Extreme Case: Augmenting Empty Graphs

A peculiar yet very interesting augmentation problem arises when the input graph has no edges: we are only given a set S of points in the plane and some graph property $\mathcal{P}$. The most restrictive graph property requires the augmentation to be isomorphic with a given planar graph $G=(V,E)$, with $|S|=|V|=n$ vertices. In other words, the problem is to *embed G with straight-line edges on top of S* without crossings.

Notice that this problem is not always feasible, and that the specific configuration of the points in S makes a world of difference. For example, no graph with at least $2n-2$ edges can be embedded on top of a set of n points in convex position, because this set admits at most $2n-3$ pairwise noncrossing geometric edges, which would constitute a triangulation of its convex hull. On the other hand, even if we fix a set of n points with a triangular convex hull (and so any triangulation has $3n-6$ edges), this point set may not accommodate all n-vertex planar graphs. An example is given in Fig. 9: Many maximal planar graphs with n vertices contain no vertex of degree 3 or less, or no vertex of degree $n-1$, and none of them can be realized on top of the set shown in Fig. 9.

Fig. 9 A set S of $n = 9$ points with a triangular convex hull (*left*). A maximal plane graph on top of S will be a triangulation T with $3n-6$ edges. However, the edges shown on the right must appear in every triangulation of S, and they only differ in the edges that triangulate the *gray* region Q. Therefore, T will have at least two vertices of degree 3, and at least one with degree $n-1$

3.1 Plane Drawings of Specific Graphs and Classes of Graphs

Given an n-vertex planar graph G and an n-element point set S, it is NP-complete to decide whether G admits a straight-line embedding on top of S. Cabello [23] proved that this is an NP-complete problem even if G is restricted to be 2-connected and 2-outerplanar.

However, there are finite point sets, perhaps much larger than n, that accommodate all n-vertex planar graphs. A point set S is called *n-universal* if every planar graph with n nodes admits a straight-line embedding on top of a subset of S. For instance, the union of vertex sets of arbitrary straight-line embeddings of *all* n-vertex planar graphs is n-universal, although this point set is quite large, with up to $e^{\Theta(n)}$ points. It is a longstanding open problem to find the smallest size of an n-universal point set for all $n \geq 0$ (see Problem 45 in [29]). It is known that an $(n-1) \times (n-1)$ section of the integer lattice is n-universal [43, 93]. However, no n-universal set of points with a subquadratic number of elements is known. Chrobak and Karloff [27] proved that every n-universal set must have at least $1.098n$ points, and Kurowski [72] improved the lower bound to $1.235n$.

It is also remarkable that for certain graph classes the configuration of the points—provided they are in general position—is no obstacle at all. For example, Gritzmann et al. [52] proved that every outerplanar graph of n vertices admits a crossing-free plane embedding on top of any set of n points in general position. There are also efficient algorithms for the construction of the embedding [15, 69]. In the case of a tree $T = (V, E)$, it is even possible to select a node $v \in V$ and a point p in the given point set S, and realize T on top of S with the additional constraint that v is mapped to p. This was proved by Ikebe et al. [63], and an efficient associated algorithm is described in [19].

There are many variants of similar embeddability problems. Brass et al. [21], for example, consider the *geometric simultaneous embedding* of two planar graphs in which two graphs should admit a plane realization on top of the same vertex set, possibly with the additional requirement that they are compatible; see [42] for a list of results on this subject. We do not pursue here this extension or the many variants

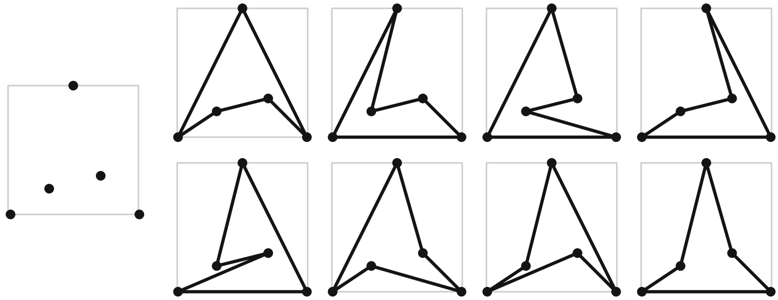


Fig. 10 A set S of five *points* in the plane, and all eight *straight-line* embeddings of C_5 on top of S

that appear in the rich literature of the *Graph Drawing* area of problems. We refer the interested reader to Chap. 9 in [22] for geometric graphs, and to [31, 101], Chap. 21 in [91], and Chap. 52 in [49] for the more generic framework.

3.2 The Number of Embeddings

Some graphs admit many straight-line embeddings on top of a point set S in general position. For instance, the cycle C_n has several embeddings on an n-element point set, unless the points are in convex position. It was the subject of intense research to find upper and lower bounds on the maximum possible number of embeddings. Here we briefly survey two cases and provide references for other graph classes.

Let S be a set of n points in the plane in general position, labeled by integers from 1 through n. There is an obvious combinatorial upper bound on the number of polygonizations given by the number of cyclic permutations, $(n-1)!$, which is roughly n^n, neglecting exponential and polynomial terms. However, it is clear that most of the polygons generated in this way would have crossings. Notice that the solution to the Euclidean traveling salesman problem, that is, the spanning cycle of minimal total length, is necessarily crossing-free, i.e., a polygonization, because any two crossing edges can be replaced by two edges of their convex hull, resulting in a strictly shorter spanning cycle. This motivates why this counting problem, going back to Newborn and Moser [78], has been intensely investigated (Fig. 10).

A major step in estimating the number straight-line embeddings was achieved by Ajtai, Chvátal, Newborn, and Szemerédi [6]. They proved that every set of n points in the plane admits at most c^n crossing-free geometric graphs, with a constant $c = 10^{13}$. It is worth mentioning that the *crossing lemma*, a cornerstone in geometric graph theory, was proved in this paper as a lemma for this result[1]. The upper bound for the number of polygonizations on an n-element point set has

[1]The crossing lemma was independently proved by Leighton [73] as well.

been improved substantially in a series of papers. One of the latest upper bounds is $O(68.66^n)$ by Dumitrescu et al. [35], which has recently been further improved to $O(54.55^n)$ by Sharir et al. [95]. The best current lower bound comes from the *double-chain* configuration (see Fig. 11), which is known to have more than 4.64^n different polygonizations [45]. If we denote by $p(n)$ the maximum number of polygonizations over all sets of n points in general position, the currently best bounds can be summarized as $4.64^n \lesssim p(n) \lesssim 54.55^n$, where we omit polynomial factors and display the dominating exponential term only.

The bound c^n given by Ajtai et al. [6] also applies to triangulations. Counting triangulations is an old problem going back to Euler, who considered the case of points in convex position, counted by the *Catalan numbers* (see [94] for a historical account). Possibly David Avis was the first to ask about the maximal number of triangulations over generic point sets [110]. Denote by $\mathrm{tr}(S)$ the number of triangulations a point set S admits, and by $\mathrm{tr}(n) = \max_{|S|=n} \mathrm{tr}(S)$ the maximal value over all n-element point sets. Assume that we have a bound $\mathrm{tr}(n) \leq c_t^n$ for some constant c_t. Then, since every crossing-free geometric graph on n vertices can be augmented to a triangulation and a triangulation has at most $3n-6$ edges, we infer that $2^{3n-6}c_t^n \approx 8^n c_t^n$ is an upper bound for the number of plane geometric graphs on top of any n-element point set.

Similarly, if, for some class of n-vertex graphs $\mathcal{G}$, one can derive an upper bound of g^n on the number of graphs in $\mathcal{G}$ that are contained as a subgraph in a triangulation on n vertices, then $g^n c_t^n$ is an upper bound for the different realizations of graphs in $\mathcal{G}$ on any n-element point set. This explains why improving on the value of the constant c_t has been the subject of a long list of papers; see [97, 110] for an account and references. Currently, the best upper bound is $\mathrm{tr}(n) \leq 30^n$, obtained by Sharir and Sheffer [94], refining the method given by Sharir and Welzl [96], which combines the use of random triangulations with a charging scheme. On the opposite direction, the double-chain point configuration admits 8^n triangulations [45]. This lower-bound construction was widely believed to be best possible until Aichholzer et al. [5] introduced the *double-zigzag chain* and proved that it admits $\Omega(8.48^n)$ triangulations (Fig. 11). The double-chain consists essentially of two "flat" convex polygons P and Q facing each other. In the double-zigzag chain, P and Q are the most basic type of *almost-convex polygons*, a class introduced by Hurtado and Noy [59]. Using other almost-convex polygons in a similar manner, Dumitrescu et al. [35] have recently constructed n-element point sets that admit $\Omega(8.65^n)$ triangulations. While the gap between the upper and lower bounds keeps narrowing and there is no firm conjecture on what the right constant may be in the exponent, nonetheless one may say that the lower bounds have always been thought to be closer to the true value.

Other graphs, such as perfect matchings, spanning paths, pseudo-triangulations, and many more have been studied from this perspective. Details on bounds and references can be found in [5, 97].

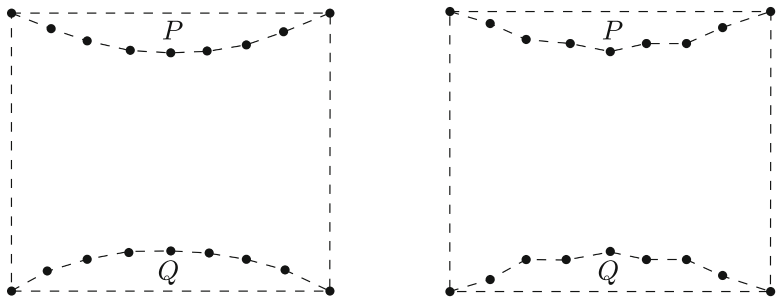


Fig. 11 The double-chain (*left*) consists of two convex chains that face each other. The line connecting any pair of points in the lower (resp., upper) chain passes below (resp., above) all the points in the other chain. The *dashed* edges are present in every triangulation, so the three resulting regions are triangulated independently. The double-zigzag chain *(right)* is very similar, but reflex and convex vertices alternate, and the *line* defined by two consecutive vertices in the lower chain leaves exactly one point from the lower chain above; the situation is symmetric for the upper chain

3.3 Spanning Graphs with Desirable Properties

In many applications, we do not necessarily want to embed a specific graph on top of a point set S, but rather we would like to construct a spanning graph on S with certain properties. The properties may be purely graph-theoretical (e.g., connectivity) or specific to the geometric realization (e.g., dependent on Euclidean lengths or angles). In this section, we consider some representative examples of each type.

Connectivity. Given a point set S in the plane in general position, and an integer $k \geq 0$, we could like to embed a k-connected (resp., k-edge-connected) graph on top of S. For $k = 1$, it is clear that every point set admits a spanning tree, which is the smallest connected graph on n vertices. As noted above, $n \geq 3$ points in general position also admit a spanning cycle, which is the smallest 2-connected graph on n vertices. However, n points in convex position do not admit 3-connected augmentation, since every maximal augmentation is a triangulation in which at least two vertices have degree 2. It is easy to see, though, that any other point set (i.e., not in convex position) on $n \geq 4$ vertices admits a 3-connected graph: Start with a star centered at a vertex in the interior of the convex hull, and complete it into a wheel graph, which is 3-connected. For 4- and 5-connectivity, the analogous question about feasibility is not so easy anymore.

To determine whether a point set admits a k-connected graph, it is enough to consider whether it admits a k-connected *triangulation*, since every planar straight-line graph can be triangulated and additional edges can only increase the connectivity. Dey et al. [30] proved that every set whose convex hull consists of 3 vertices admits a 4-connected triangulation, with the only exception of the point set

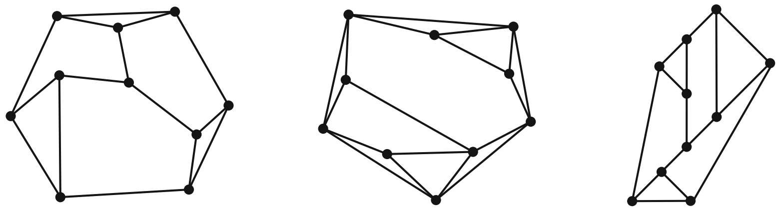


Fig. 12 A set of $n = 10$ points, $h = 6$ of which lie on the boundary of the convex hull, admits a 3-connected cubic PSLG (*left*). A minimum convex decomposition for $n = 10$ points, $h = 5$, with $\lfloor \frac{7}{3}n - h \rfloor = 18$ edges (*middle*). A *straight-line* embedding of a 3-connected cubic graph where the 15 edges have only 6 different slopes (*right*)

in Fig. 9, left. No characterization is known for generic point sets (with arbitrary convex hull) that admit 4- or 5-connected triangulations. Note that k-connectivity is infeasible for $k \geq 6$, since very planar graph has a vertex of degree at most 5.

If we can decide whether there exists a k-connected graph on top of a given point set S, the next question is to find one with the fewest possible edges. If we allow crossings, then every set of n vertices admits a 3-connected graph with $\lceil 3n/2 \rceil$ edges, which is the best possible since the degree of every vertex must be at least 3. For constructing a *crossing-free* 3-connected graph on top of a point set S, the location of the points (or, rather, their order type) already matters. García et al. [44] proved that if we are given a set of n points, $h < n$ of which lie on the boundary of the convex hull, then it admits a 3-connected planar straight-line graph with $\max(\lceil 3n/2 \rceil, n+h-1)$ edges, and this bound is the best possible for each point set. This implies, in particular, that a cubic (that is, absolute minimum size) 3-connected graph is possible if $n \geq 4$ is even and at most $n/2+1$ points lie on the convex hull (Fig. 12, left). The proof in [44] is algorithmic, and 3-connected graphs of the above size can be computed in polynomial time.

Angle and Slope Conditions. A frequently used generalization of triangulations is a *convex decomposition*, which is a plane geometric graph where all bounded faces are convex, and the bounded faces jointly tile the convex hull of the vertex set. Every point set admits a convex decomposition (since every triangulation is a convex decomposition). While a triangulation on top of n points, h of which lie on the convex hull, has exactly $3n-h-3$ edges, a convex decomposition may have as few as n edges (if the points are in convex position). Improving upon earlier results [58, 77], Sakai and Urrutia [92] showed recently that every set of n points, h of which are on the convex hull, admits a convex decomposition with $\frac{7}{3}n - h$ edges, which is the best possible for infinitely many (n,h) pairs (Fig. 12, middle), but not known to be optimal in general. The currently best lower bound, $\frac{23}{11}n - 3$, in terms of n is due to García-Lopez and Nicolás [46].

It is well known [12] that among all triangulations of a point set, the Delaunay triangulation maximizes the minimum angle (intuitively, all triangles are as "fat" as possible). Aichholzer et al. [4] proved that every point set admits a triangulation in which each point is incident to a triangle that has an angle of at least $2\pi/3$ at that point, and the bound $2\pi/3$ is the best possible. There are similar results [4,14,34,39] for angle-constrained spanning cycles; however, all known results consider arbitrary geometric graphs (*with* possible crossings). It is an open problem to determine the minimum and maximum possible angles that a *crossing-free* spanning cycle, spanning path, or spanning tree can have, over all n-element point sets in general position.

In a problem closely related to the angle constraints on adjacent edges, Dujmović et al. [33] studied the crossing-free geometric graphs that can be embedded in the plane with few different edge *slopes*. They prove that every planar graph with n vertices has a straight-line embedding in the plane with at most $2n-10$ different slopes, based on a canonical decomposition introduced by de Fraysseix [43]. They also construct triangulations that require at least $n+2$ different slopes. It remains an open problem to determine the minimum number of slopes sufficient for the straight-line embedding of every planar graph with n vertices. It is clear that a vertex of degree Δ forces at least $\lceil \Delta/2 \rceil$ slopes in any straight-line drawing. Dujmović et al. [33] show that $\lceil \Delta/2 \rceil$ slopes are sufficient for the embedding of every tree of maximum degree Δ; and six slopes are sufficient for the embedding of every 3-connected cubic planar graph [67] (Fig. 12, right). But, in general, it is not known whether the "slope number" of a planar graph is a function of the maximum degree.

Spanners. The *stretch factor* of a geometric graph $G=(V,E)$ is the maximum ratio

$$\max_{u,v\in V} \frac{|\mathrm{path}(u,v)|}{|uv|},$$

where $|\mathrm{path}(u,v)|$ is the Euclidean length of the shortest path between u and v, and $|uv|$ is the Euclidean distance between points u and v. The ratio $|\mathrm{path}(u,v)|/|uv|$ is also called the *detour* between points p and q. A geometric graph $G=(V,E)$ is a k-spanner if its stretch factor is at most k. For a set of n points, the only 1-spanner is the complete geometric graph, which has $\binom{n}{2}$ edges and many crossings whenever $n \geq 5$. Chew [26] proved that for every point set, there is a plane geometric graph with stretch factor at most 2, and it is the dual graph of the Voronoï diagram induced by equilateral triangles. Chew conjectured that the stretch factor of the standard Delaunay triangulation (induced by disks) is at most $\pi/2 \approx 1.5708$. Chew's conjecture has recently been refuted by Bose et al. [17]. The best current lower bound was found by a computer search: Xia and Zhang [113] presented a point set whose Delaunay triangulation has a stretch factor of 1.5907. The best current upper bound, 1.998, has recently been announced by Xia [112], improving the previous best upper bound of $4\sqrt{3}\pi/9 \approx 2.418$ by Keil and Gutwin [70]. It is not known what the maximum stretch factor of a Delaunay triangulation is, and, more interestingly, the minimum stretch factor that a plane geometric graph can have on top of any finite point set.

Giannopoulos et al. [48] and Gudmundsson and Smid [50] proved independently that, for a given set of n points and an integer K, it is NP-hard to find the plane geometric graph of minimum stretch factor and at most K edges. It is still possible, however, that one can efficiently compute a plane geometric graph with minimum stretch factor for a given point set if there is no limitation on the number of edges.

In the last 20 years, researchers have extensively studied crossing-free spanners that have several desirable properties simultaneously, such as bounded stretch factor, bounded degree, and small Euclidean length [16]. One recent result in this thread, by Kanj et al. [66], states that for every integer $k \geq 14$ and every n-element point set, there is a PSLG with maximum degree at most k and stretch factor at most $(1 + \frac{2\pi}{k\cos(\pi/k)}) \cdot 4\sqrt{3}\pi/9$, and such a PSLG can be computed in $O(n)$ time. For a detailed historical account, variants of the problem, and algorithmic results, we refer to a survey book by Narasimhan and Smid [76].

4 Generic Plane Augmentation Problems

In the previous two sections, we surveyed augmentation problems for perfect matchings and empty graphs. In the most general version of geometric graph augmentation, we are given an arbitrary planar straight-line graph $G = (V, E)$ and a graph property $\mathcal{P}$. Our task is to augment G into a PSLG with property $\mathcal{P}$ or report that no such augmentation is possible. Here $\mathcal{P}$ may be an abstract graph property or a property specific to geometric graphs. At any rate, we can consider only properties of planar graphs, since the output has to be planar. We consider two illustrative examples: One is connectivity augmentation, where property $\mathcal{P}$ is k-connectivity or k-edge-connectivity for some integer k, $1 \leq k \leq 5$. The other property is *detour at most* d, for some real $d \geq 1$, which is specific to geometric graphs.

4.1 Connectivity Augmentation

The k-connectivity (resp., k-edge-connectivity) augmentation problem asks for the minimum number of edges that augment a given graph $G = (V, E)$ into a k-connected (resp., k-edge-connected) graph $G' = (V, E \cup E')$. For abstract graphs, both the vertex- [65] and edge-connectivity [41, 75, 109] versions have polynomial-time solutions for every integer k; and there are linear-time solutions for $k = 2$ [37, 60, 61, 86]. In general, vertex connectivity is technically more difficult to handle. When k is part of the input, Végh [108] gave a polynomial-time algorithm for the k-connectivity augmentation of a $(k-1)$-connected graph.

Kant and Bodlaender [68] considered connectivity augmentation problems over planar graphs, where both the input and output are required to be planar (i.e., no minor isomorphic to K_5 or $K_{3,3}$). They showed that such a *planarity-preserving* edge-connectivity augmentation is NP-hard already for $k = 2$, and later Rutter and Wolff [90] showed this for vertex connectivity as well. Fialko and Mutzel [40] and Gutwenger et al. [51] proposed constant-factor approximations. However, in a

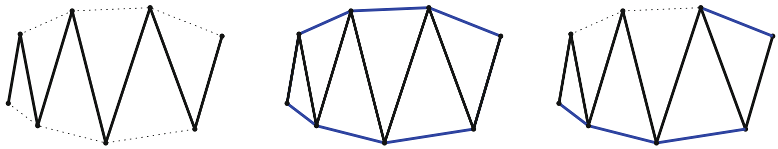


Fig. 13 A zigzag path on eight vertices in convex position (*left*). Augmentation to 2connectivity requires six new edges (*middle*). Augmentation to 2-edge connectivity requires four new edges (*right*)

planarity-preserving augmentation, the output graph may not be compatible with every straight-line embedding of the input graph, and so these results are not applicable when the input is a planar straight-line graph.

Connectivity augmentation over planar straight-line graphs is significantly more restrictive. For instance, a path on n vertices has a planarity-preserving augmentation to a 2-connected Hamiltonian cycle by adding one edge (between the endpoints of the path). However, if the path is embedded in the plane as a zigzag chain on n points in convex position (Fig. 13), then 2-edge-connectivity augmentation requires $\lfloor n/2 \rfloor$ new edges, 2-connectivity requires $n-2$ new edges, and the connectivity augmentation is not even feasible for $k \geq 3$. Rutter and Wolff [90] proved that the k-connectivity and k-edge-connectivity augmentation problems are NP-hard over plane geometric graphs for $k = 2, 3, 4$, and 5. As noted above, the problem is infeasible for $k \geq 6$ because every planar graph has a vertex of degree at most 5. Currently, there are no approximation algorithms. Research focused on determining whether augmentation is feasible and on extremal bounds: What is the minimum number of new edges that are sufficient for the k-connectivity (resp., k-edge-connectivity) augmentation of *any* plane geometric graph on n vertices?

Vertex Connectivity. Abellanas et al. [1] were the first to obtain extremal bounds. They proved that a connected PSLG with b cut vertices can be augmented to a 2-connected PSLG with b new edges. The case of the zigzag path, with $n-2$ cut vertices, shows that this bound is the best possible. It follows that every connected PSLG on n vertices can be augmented to a 2-connected PSLG with at most $n-2$ new edges, and this bound is also the best possible. It is not known, however, what is the minimum number of edges sufficient for the 2-connectivity augmentation of any (possibly disconnected) PSLG with n vertices.

It is easy to see that every PSLG in general position can be augmented to 2-connectivity, since every triangulation is 2-connected. This is no longer true for 3-connectivity. Tóth and Valtr [105] proved that a PSLG $G = (V, E)$ can be augmented to 3-connectivity if and only if V is not in convex position and no edge in E is a proper diagonal of the convex hull $\mathrm{CH}(V)$. Similar combinatorial characterizations are not known for 4- or 5-connectivity.

Edge Connectivity. Every PSLG can be augmented to a 2-edge-connected PSLG. Tóth and Valtr [105] characterized the PSLGs that can be augmented to 3-edge connectivity; they are called *3-edge-augmentable*. Specifically, a PSLG $G = (V, E)$

Table 1 The minimum number of new edges sufficient for raising the edge connectivity of any PSLG on n vertices in general position to a target $k = 1, 2, 3$ if augmentation is possible. Tight bounds and lower bounds

Target edge connectivity	**1**	**2**	**3**
Arbitrary PSLG	$n-1$	$\geq \lfloor (4n-4)/3 \rfloor$	$2n-2$
Connected PSLG	0	$\lfloor (2n-2)/3 \rfloor$	$\geq \lfloor (4n-4)/3 \rfloor$
2-edge-connected PSLG	0	0	$n-2$

is 3-edge-augmentable if and only if V is not in convex position and there is no edge $e \in E$ such that e is a proper diagonal of $\mathrm{CH}(V)$ and all vertices on one side of e lie on the boundary of $\mathrm{CH}(V)$. Similar combinatorial characterizations are not known for 4- or 5-connectivity.

Let us point out two easy but important tools that simplify the study of edge-connectivity augmentation. The first tool is the concept of *k-edge-connected* components in a graph $G = (V, E)$: It is a maximal subset $V_c \subset V$ such there are at least k edge-disjoint paths between any two vertices of V_c. By Menger's theorem, a graph G is k-edge-connected if and only if it has only one k-edge-connected component. The second tool is the use of multiedges. Abellanas et al. [1] proved that if a PSLG $G = (V, E)$ can be augmented to 2-edge connectivity with m new edges such that some new edges are duplications of existing edges in E, then the augmentation is also possible with m new edges and no double edges. An analogous result holds for 3-edge-connectivity augmentation if G is 3-edge-augmentable [8].

Extremal bounds are known for the minimum number of edges sufficient for increasing the edge connectivity by one for any PSLG with n vertices (see Table 1). It is easy to see that every disconnected PSLG with $c \geq 2$ components can be augmented to a connected PSLG by adding exactly $c-1$ new edges. The maximum value of c is n, attained for empty graph. Abellanas et al. [1] conjectured that every connected PSLG on n vertices can be augmented to a 2-edge-connected PSLG with at most $\lfloor (2n-2)/3 \rfloor$ new edges, and the this was later confirmed by Tóth [104]. This upper bound is the best possible: The lower-bound construction consists of a triangulation with m vertices, with a leaf added in each of the $2m-5$ bounded faces, and three pairwise invisible leaves in the outer face (Fig. 14, left). This PSLG has $n = m + (2m-5) + 3 = 3m-2$ vertices, and each of the $2m-2$ leaves requires one new edge for 2-edge connectivity. The proof of the upper bound is constructive: A polynomial-time augmentation algorithm adds new edges in each face of the input graph independently. The key tool is the *geodesic hull* of all 2-edge-connected components adjacent to a face (Fig. 14, middle). The boundary of the geodesic hull is a closed polygonal chain P that visits all 2-edge-connected components, and so one can add some edges of P into the input graph (possibly creating double edges), and merge them into a single 2-edge-connected component (Fig. 14, middle). At the end of the algorithm, double edges can be replaced with single edges by [1].

Tóth and Valtr [105] showed that $n-2$ edges are always sufficient for raising the edge connectivity of a 3-edge-augmentable PSLG on n vertices from 2 to 3. The upper bound is based on the simple observation that a new edge can decrease the

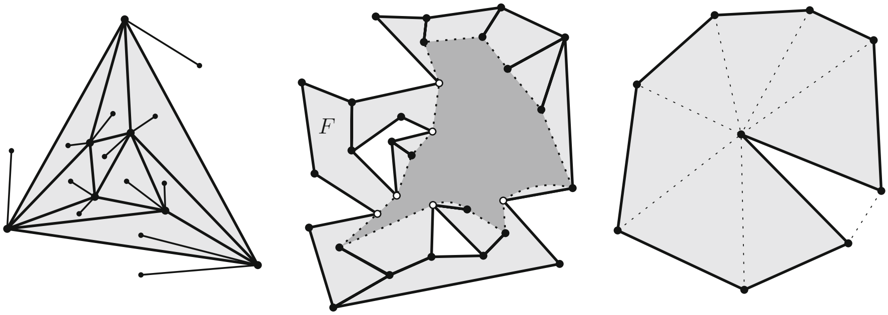


Fig. 14 APSLG on $n = 19$ vertices, including $\lfloor (2n-2)/3 \rfloor = 12$ leaves (*left*). The geodesic hull of one vertex of each 2-edge-connected component incident to a face F (*middle*). A 2-edge-connected graph on $n = 9$ vertices that requires $n-2 = 7$ new edges to 3-edge-connectivity (*right*)

number of 3-edge-connected components by one, and if there are n such components initially, then the first two new edges can create a K_4 minor, which is a 3-connected component with 4 vertices. This upper bound is the best possible: If $G = (V,E)$ is a Hamiltonian cycle with exactly one vertex in the interior of $\mathrm{CH}(V)$, then the only 3-edge-connected augmentation is the wheel graph with $n-2$ new edges (Fig. 14, right).

For abstract graphs, one can solve k-connectivity augmentation optimally in a multiphase algorithm, where each phase increments the edge connectivity by one [25]. This is not the case for geometric graphs, since edges added in one phase could become obstacles in subsequent phases. Al-Jubeh et al. [8] showed that any PSLG on n vertices in general position can be augmented to 3-edge connectivity with at most $2n-2$ new edges, if the augmentation is feasible, and this bound is the best possible. No tight extremal bounds are known for raising the edge connectivity from 0 to 2 or from 1 to 3 (see Table 1). Lower-bound constructions, similar to one in the left part of Fig. 14, show that in the worst case, $\lfloor (4n-4)/3 \rfloor$ new edges are needed for both problems.

Augmenting Crossing-Free Straight-Line Trees. Abellanas et al. [1] proved that every path on n vertices in general position can be augmented to 2-edge connectivity with $\lfloor n/2 \rfloor$ edges, and this bound is the best possible (Fig. 13). They conjectured that $\frac{n}{2} + O(1)$ edges are sufficient for augmenting any tree on n vertices to 2-edge connectivity. However, Tóth [104] constructed a family of trees that require $\frac{17}{33}n - O(1)$ new edges. García and Tejel (private communication) improved the lower bound to $\frac{6}{11}n - O(1)$. Consider the tree in the left part of Fig. 15. It has eight leaves, five of which lie in the interior of the convex hull and are pairwise invisible to each other. Identify one of the interior leaves A with an exterior leaf of a scaled copy of the original construction as in the right part of Fig. 15(right). Then the number of vertices goes up by 11. The leaf A, which required one new edge, has been replaced by a subgraph with seven leaves. Two of these leaves see each other, and the remaining five leaves are isolated. However, even if we join the two new

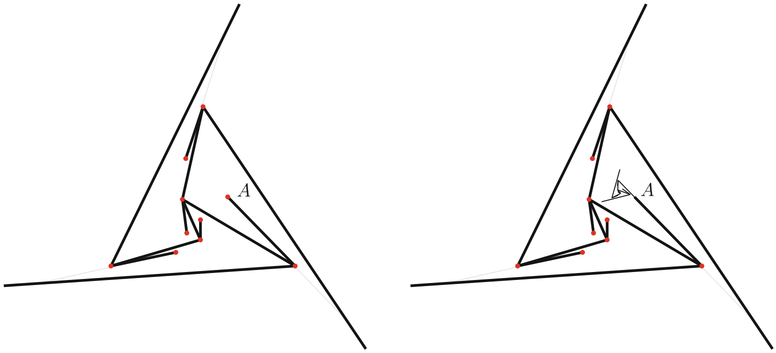


Fig. 15 A tree with 13 vertices, including 8 leaves, which requires 7 new edges (*left*). Replacing A with a scaled copy of the original construction, the number of vertices increases by 11, and the number of leaves increases by 7, and it requires 6 more new edges (*right*)

mutually visible leaves to each other, the original edge A is still not contained in any circuit, so these seven leaves require seven new edges. By iterating this step k times, we obtain a tree with $13 + 11k$ vertices that requires $7 + 6k$ new edges.

Currently, the best upper bound for trees is the same as for any PSLG: Every connected PSLG with n vertices can be augmented to a 2-edge-connected PSLG by adding at most $\lfloor (2n-2)/3 \rfloor$ new edges. We emphasize again that the difficulty is posed by requiring *straight-line* edges. If the new edges are allowed to be Jordan arcs, then the lower bound of $\lfloor n/2 \rfloor$ is the best possible [104].

4.2 Network Optimization Through Augmentation

Motivated by wireless communication networks, Kranakis et al. [71] studied 2-edge-connectivity augmentation for noncrossing subgraphs of unit disk graphs. They prove that if $G = (V, E)$ is a connected PSLG in general position, it has b cut edges, and all edges have length at most 1, then G can be augmented to a 2-edge-connected PSLG with at most b new edges, each of which has length at most 3. The bound 3 on the length of the new edges cannot be improved (Fig. 16, left), but the bound on the number of new edges is not known to be optimal. Similarly, a noncrossing tree with n vertices in general position and edges of length at most 1 can be augmented to a 2-edge-connected PSLG with at most $\lfloor 5n/6 \rfloor$ new edges, each of length at most 3. It follows from the reduction of Rutter and Wolff [90] that finding the minimum number of new edges remains NP-complete under the restriction that the new edges have bounded length.

In typical applications in wireless communication, the edges of a graph or digraph are given implicitly. The input is a set S of points in the plane. Every assignment of angular domains with apices at S and radius r (which correspond to

Fig. 16 A path with unit length edges, which requires an edge of length $3-\varepsilon$ for 2-edge-connectivity augmentation (*left*). A path between a and b, where the stretch factor is large but cannot be improved with additional edges (*right*)

directional antennae of uniform range r stationed at points in S) defines a geometric digraph $G=(V,E)$ such that $(u,v)\in E$ if and only if one of the angular domains with apex at u contains v and the distance $|uv|$ is at most r. A digraph defined in this way can be augmented by increasing the radius r. Note, however, that some edge pairs may inevitably cross in this model. Problems in the literature [24, 28, 32] ask for the minimum radius and minimum angle (alternatively, sum of angles) that guarantee strong connectivity.

Giannopoulos et al.[48] showed that it is NP-hard to decide whether a given PSLG $G=(V,E)$ can be augmented with at most K new edges such that its stretch factor drops below a given threshold $\lambda>1$. Farshi et al. [38] considered the problem of finding a single new edge that would maximally decrease the stretch factor of a geometric graph, where both the input graph and the new edge may have crossings; they proposed efficient algorithms for both finding and approximating the optimal "shortcut." Wulff-Nilsen [111] generalized the problem to arbitrary metric spaces. Aronov et al. [10] studied the variant of this problem where the input graph contains a singleton (which must be connected to the remainder of the graph), but they also allow crossings in the output. No results are known for finding such "bottleneck" edges over crossing-free geometric graphs in the plane. Note that if crossings are not allowed, then the stretch factor of the augmentation cannot be decreased below any constant threshold. The detour between two endpoints of a zigzag path can be arbitrarily large (Fig. 16, right), but in some cases it cannot be improved by adding new edges.

In all applications mentioned so far, it was clear that the addition of new edges can only improve the network (new edges can neither decrease the connectivity nor increase the detour). We note here that augmentation may deteriorate the quality of a network. For example, there are simple instances of noncooperative network congestion games (cf. Braess's paradox) where the total latency at Nash equilibrium increases if we add a new edge. Roughgarten [88, 89] showed that this deterioration can be arbitrarily large, even in planar networks. Network design problems, under various capacity and demand constraints, go beyond the scope of this chapter. We refer to a comprehensive survey by Tardos and Wexler [102]. Current techniques in this domain are unable to enforce the crossing-free condition, and it would be very interesting to study congestion games for embedded networks in a geometric setting.

5 Concluding Remarks

The family of augmentation problems is quite large. As mentioned in the Introduction, we have only discussed some selected representatives problems here. We conclude with briefly mentioning some problems that have not been included.

- We omitted the *weighted versions* of problems, in which the weight of an edge is its Euclidean length, and the weight of a PSLG is its total edge length. The complexity of only very few weighted optimization problems are known. Given a set of n points in the plane, the Euclidean minimum spanning tree (EMST) can be computed in $O(n \log n)$ time, as a minimum spanning tree in the Delaunay triangulation of the points. However, finding a minimum-weight triangulation is NP-hard [79]. The minimum-weight PSLGs from most graph classes (*e.g.*, spanning trees, spanning cycles) are often automatically noncrossing, but the crossing-free property has to be explicitly enforced for *maximum*-weight PSLGs, which is impossible to model with standard optimization techniques. Approximation algorithms for some variants have been presented in [9, 36].
- We have not reviewed *combinatorial games*, in which two players construct disjoint compatible PSLGs on a given point set by incrementally augmenting the empty graph, following certain rules. Many graph creation games have been considered (e.g., maker-breaker or avoider-enforcer games) for noncrossing geometric graphs [3], but many open problems remain.
- We did not mention any problem that involves Steiner points (that is, augmenting a graph with both vertices and edges, and possibly subdividing existing edges). Patrignani [83] showed that it is NP-complete to decide whether a partial embedding of a graph can be completed to a straight-line embedding. This problem is, in fact, NP-complete already for trees [13].

Acknowledgements Research by Ferran Hurtado has been partially supported by projects MEC MTM2009-07242, ESF EUROCORES programme EuroGIGA, CRP ComPoSe: MICINN Project EUI-EURC-2011-4306, and Gen. Catalunya DGR 2009SGR1140. Research by Csaba Tóth has been partially supported by NSERC Grant RGPIN 35586 and the NSF Grant CCF-0830734, by the Tóth has been conducted at Tufts University, Medford, MA, USA.

References

1. M. Abellanas, A. García, F. Hurtado, J. Tejel, J. Urrutia, Augmenting the connectivity of geometric graphs. Comput. Geom. Theor. Appl. **40**(3), 220–230 (2008)
2. O. Aichholzer, S. Bereg, A. Dumitrescu, A. García, C. Huemer, F. Hurtado, M. Kano, A. Márquez, D. Rappaport, S. Smorodinsky, D. Souvaine, J. Urrutia, D. Wood, Compatible geometric matchings. Comput. Geom. Theor. Appl. **42**, 617–626 (2009)
3. O. Aichholzer, D. Bremner, E.D. Demaine, F. Hurtado, E. Kranakis, H. Krasser, S. Ramaswami, S. Sethia, J. Urrutia, Games on triangulations. Theor. Comp. Sci. **343**(1–2), 42–71 (2005)

4. O. Aichholzer, T. Hackl, M. Hoffmann, C. Huemer, A. Pór, F. Santos, B. Speckmann, B. Vogtenhuber, Maximizing maximal angles for plane straight line graphs, in *Proceedings of the 10th Workshop on Algorithms and Data Structures*, LNCS, vol. 4619 (Springer, Berlin, 2007), pp. 458–469
5. O. Aichholzer, T. Hackl, C. Huemer, F. Hurtado, H. Krasser, B. Vogtenhuber, On the number of plane geometric graphs. Graphs Combinator. **23**(1), 67–84 (2007)
6. M. Ajtai, V. Chvátal, M.M. Newborn, E. Szemerédi, Crossing-free subgraphs. Ann. Discrete Math. **12**, 9–12 (1982)
7. M. Al-Jubeh, M. Hoffmann, M. Ishaque, D.L. Souvaine, C.D. Tóth, Convex partitions with 2-edge connected dual graphs. J. Combin. Opt., **22**(3), 409–425 (2011)
8. M. Al-Jubeh, M. Ishaque, K. Rédei, D.L. Souvaine, C.D. Tóth, Tri-edge-connectivity augmentation for planar straight line graphs. Algorithmica **61**(4), 971–999 (2011)
9. N. Alon, S. Rajagopalan, S. Suri, Long non-crossing configurations in the plane. Fundam. Inform. **22**(4), 385–394 (1995)
10. B. Aronov, K. Buchin, M. Buchin, M. Van Kreveld, M. Löffler, J. Luo, R.I. Silveira, B. Speckmann, Connect the dot: computing feed-links with minimum dilation, in *Proceedings of the Algorithms Data Structures Symposium (WADS)*, LNCS, vol. 5664 (Springer, Berlin, 2009), pp. 49–60
11. T. Asano, S.K. Ghosh, T.C. Shermer, in *Visibility in the Plane*, ed. by J.R. Sack, J. Urrutia. Handbook on Computational Geometry (Elsevier, Amsterdam, 2000), pp. 829–876
12. F. Aurenhammer, Y. Xu, in *Optimal Triangulations*, ed. by C.A. Floudas, P.M. Pardalos. Encyclopedia of Optimization, vol. 4 (Kluwer, Dordrecht, 2001), pp. 160–166
13. A. Bagheri, M. Razzazi, Planar straight-line point-set embedding of trees with partial embeddings. Inf. Proc. Lett. **110**, 521–523 (2010)
14. I. Bárány, A. Pór, P. Valtr, Paths with no small angles. SIAM J. Discrete Math. **23**(4), 1655–1666 (2009)
15. P. Bose, On embedding an outer-planar graph in a point set. Comput. Geom. Theor. Appl. **23**, 303–312 (2002)
16. P. Bose, J. Gudmundsson, M. Smid, Constructing plane spanners of bounded degree and low weight. Algorithmica **42**, 249–264 (2005)
17. P. Bose, L. Devroye, M. Löffler, J. Snoeyink, V. Verma, The spanning ratio of the Delaunay triangulation is greater than $\pi/2$. Comput. Geom. Theor. Appl. **44**(2), 121–127 (2011)
18. P. Bose, M.E. Houle, G.T. Toussaint, Every set of disjoint line segments admits a binary tree. Discrete Comput. Geom. **26**(3), 387–410 (2001)
19. P. Bose, M. McAllister, J. Snoeyink, Optimal algorithms to embed trees in a point set. J. Graph Algorithm. Appl. **1**(2), 1–15 (1997)
20. P. Bose, G.T. Toussaint, Growing a tree from its branches. J. Algorithm. **19**, 86–103 (1995)
21. P. Brass, E. Cenek, C. Duncan, A. Efrat, C. Erten, D. Ismailescu, S. Kobourov, A. Lubiw, J. Mitchell. On simultaneous planar graph embeddings. Comput. Geom. Theor. Appl. **36**(2), 117–130 (2007)
22. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry*. (Springer, Berlin, 2005)
23. S. Cabello, Planar embeddability of the vertices of a graph using a fixed point set is NP-hard. J. Graph Algorith. Appl. **10**, 353–363 (2006)
24. I. Caragiannis, C. Kaklamanis, E. Kranakis, D. Krizanc, A. Wiese, Communication in wireless networks with directional antennas, in *Proceedings of the 20th ACM Symposium on Parallel Algorithms and Architectures* (ACM, New York, 2008), pp. 344–351
25. E. Cheng, T. Jordán, Succesive edge-connectivity augmentation problems. Math. Program. **84**, 577–593 (1999)
26. L.P. Chew, There are planar graphs almost as good as the complete graph. J. Comput. Syst. Sci. **39**(2), 205–219 (1989)
27. M. Chrobak, H. Karloff, A lower bound on the size of universal sets for planar graphs. SIGACT News **20**, 83–86 (1989)

28. M. Damian, R. Flatland, Spanning properties of graphs induced by directional antennas, in *Proceedings 20th Fall Workshop on Computational Geometry* (Stony Brook, NY, 2010)
29. E.D. Demaine, J.S.B. Mitchell, J. O'Rourke (eds.) *The Open Problems Project.* http://cs.smith.edu/$^\sim$orourke/TOPP/welcome.html
30. T.K. Dey, M.B. Dillencourt, S.K. Ghosh, J.M. Cahill, Triangulating with high connectivity. Comput. Geom. Theor. Appl. **8**, 39–56 (1997)
31. G. Di Battista, P. Eades, R. Tamassia, I.G. Tollis, *Graph Drawing: Algorithms for the Visualization of Graphs* (Prentice Hall, Englewood Cliffs, NJ, 1998)
32. S. Dobrev, E. Kranakis, D. Krizanc, O. Morales, J. Opatrny, L. Stacho, Strong connectivity in sensor networks with given number of directional antennae of bounded angle, in *Proceedings of the 4th Conference on Combinatorial Optimization and Applications*, LNCS, vol. 6509 (Springer, Berlin, 2010), pp. 72–86
33. V. Dujmović, D. Eppstein, M. Suderman, D.R. Wood, Drawings of planar graphs with few slopes and segments. Comput. Geom. Theor. Appl. **38**, 194–212 (2007)
34. A. Dumitrescu, J. Pach, G. Tóth, Drawing Hamiltonian cycles with no large angles, in *Proceedings of the 17th Symposium on Graph Drawing (GD'09)*, LNCS, vol. 5849 (Springer, Berlin, 2010), pp. 3–14
35. A. Dumitrescu, A. Schulz, A. Sheffer, C.D. Tóth, Bounds on the maximum multiplicity of some common geometric graphs, in *Proceedings of the Symposium on Theorerical Aspects of Computer Science (STACS)*, LIPICS, Schloss Dagstuhl, vol. 5 of LIPICS (Schloss Dagstuhl, Germany, 2011), pp. 637–648
36. A. Dumitrescu, C.D. Tóth, Long non-crossing configurations in the plane. Discrete Comput. Geom. **44**(4), 727–752 (2010)
37. K.P. Eswaran, R.E. Tarjan, Augmentation problems. SIAM J. Comput. **5**(4), 653–665 (1976)
38. M. Farshi, P. Giannopoulos, J. Gudmundsson, Improving the stretch factor of a geometric network by edge augmentation. SIAM J. Comput. **38**(1), 226–240 (2008)
39. S.P. Fekete, G.J. Woeginger, Angle-restricted tours in the plane. Comput. Geom. Theor. Appl. **8**(4), 195–218 (1997)
40. S. Fialko, T.P. Mutzel, A new approximation algorithm for the planar augmentation problem, in *Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithms*, ACM Press. 1998, pp. 260–269
41. A. Frank, Augmenting graphs to meet edge-connectivity requirements. SIAM J. Discrete Math. **5**(1), 22–53 (1992)
42. F. Frati, M. Kaufmann, S. Kobourov, Constrained simultaneous and near-simultaneous embeddings, in *Proceedings 15th Symposium on Graph Drawing (GD'07)*, LNCS, vol. 4875 (Springer, Berlin, 2008), pp. 268–279
43. H. de Fraysseix, J. Pach, R. Pollack, How to draw a planar graph on a grid. Combinatorica **10**(1), 41–51 (1990)
44. A. García, C. Huemer, F. Hurtado, J. Tejel, P. Valtr, On triconnected and cubic plane graphs on given point sets. Comput. Geom. Theor. Appl. **42**(9), 913–922 (2009)
45. A. García, M. Noy, J. Tejel, Lower bounds on the number of crossing-free subgraphs of K_N. Comput. Geom. Theor. Appl. **16**, 211–221 (2000)
46. J. García-Lopez, M. Nicolás, Planar point sets with large minimum convex partitions, in *Abstracts of the 22nd European Workshop on Computational Geometry* (Delphi, Greece, 2006), pp. 51–54
47. S.K. Ghosh, *Visibility Algorithms in the Plane* (Cambridge University Press, Cambridge, 2007)
48. P. Giannopoulos, R. Klein, C. Knauer, et al., Computing geometric minimum-dilation graphs is NP-hard. Int. J. Comput. Geom. Appl. **20**(2), 147–173 (2010)
49. J.E. Goodman, J. O'Rourke (eds.), *Handbook of Discrete and Computational Geometry*, 2nd edn. (Chapman & Hall and CRC Press, Boca Raton, FL, 2004)
50. J. Gudmundsson, M. Smid, On spanners of geometric graphs. Int. J. Found. Comp. Sci. **20**(1), 135–149 (2009)

51. C. Gutwenger, P. Mutzel, B. Zey, On the hardness and approximability of planar biconnectivity augmentation, in *Proceedings of the 15th Computing and Combinatorics Conference (COCOON)*, LNCS, vol. 5609 (Springer, Berlin, 2009), pp. 249–257
52. P. Gritzmann, B. Mohar, J. Pach, R. Pollack, Embedding a planar triangulation with vertices at specified points. Am. Math. Monthly **98**(2), 165–166 (1991)
53. M. Hoffmann, B. Speckmann, C.D. Tóth, Pointed binary encompassing trees: simple and optimal. Comput. Geom. Theor. Appl. **43**(1), 35–41 (2010)
54. M. Hoffmann, C.D. Tóth, Alternating paths through disjoint line segments. Inf. Proc. Lett. **87**(6), 287–294 (2003)
55. M. Hoffmann, C.D. Tóth, Pointed and colored binary encompassing trees, in *Proceedings of the 21st Symposium on Computational Geometry* (ACM, New York, 2005), pp. 81–90
56. M. Hoffmann, C.D. Tóth, Segment endpoint visibility graphs are Hamiltonian. Comput. Geom. Theor. Appl. **26**(1), 47–68 (2003)
57. M. Hoffmann, C.D. Tóth, Vertex-colored encompassing graphs, manuscript (2010) http://math.ucalgary.ca/~cdtoth/colored-encom.pdf
58. K. Hosono, On convex decompositions of a planar point set. Discrete Math. **309**, 1714–1717 (2009)
59. F. Hurtado, M. Noy, Counting triangulations of almost-convex polygons. Ars Combinatoria **45**, 169–179 (1997)
60. T.-S. Hsu, Simpler and faster biconnectivity augmentation. J. Algorithm. **45**(1), 55–71 (2002)
61. T.-S. Hsu, V. Ramachandran, On finding a minimum augmentation to biconnect a graph. SIAM J. Comput. **22**(5), 889–912 (1993)
62. F. Hurtado, M. Kano, D. Rappaport, C.D. Tóth, Encompassing colored crossing-free geometric graphs. Comput. Geom. Theor. Appl. **39**(1), 14–23 (2008)
63. Y. Ikebe, M.A. Perles, A. Tamura, S. Tokunaga, The rooted tree embedding problem into points in the plane. Discrete Comput. Geom. **11**, 51–63 (1994)
64. M. Ishaque, D.L. Souvaine, C.D. Tóth, Disjoint compatible geometric matchings, in *Proceedings of the 27th Symposium on Computational Geometry (Paris, 2011)* (ACM, New York, 2011), pp. 125–134
65. B. Jackson, T. Jordán, Independence free graphs and vertex connectivity augmentation. J. Combin. Theor. Ser. B **94**, 31–77 (2005)
66. I.A. Kanj, L. Perkovic, G. Xia, On spanners and lightweight spanners of geometric graphs. SIAM J. Comput. **39**(6), 2132–2161 (2010)
67. G. Kant, Hexagonal grid drawings, in *Proceedings of the 18th Workshop in Graph-Theoretic Concepts in Computer Science (WG'92)*, LNCS, vol. 657 (Springer-Verlag, Berlin, 1993), pp. 263–276
68. G. Kant, H.L. Bodlaender, Planar graph augmentation problems, in *Proceedings of the 2nd Workshop on Algorithms and Data Structures*, vol. 519 of LNCS (Springer-Verlag, Berlin, 1991), pp. 286–298
69. M. Kaufmann, R. Wiese, Embedding vertices at points: few bends suffice for planar graphs. J. Graph Algorithm. Appl. **6**(1), 115–129 (2002)
70. J.M. Keil, C.A. Gutwin, Classes of graphs which approximate the complete Euclidean graph. Discrete Comput. Geom. **7**, 13–28 (1992)
71. E. Kranakis, D. Krizanc, O. Morales, L. Stacho, Bounded length, 2-edge augmentation of geometric planar graphs, in *Proceedings of the 4th Conference on Combinatorial Optimization and Applications*, LNCS, vol. 6508 (Springer, Berlin, 2010), pp. 385–397
72. M. Kurowski, A 1.235 lower bound on the number of points needed to draw all n-vertex planar graphs. Inf. Proc. Lett. **92**, 95–98 (2004)
73. T. Leighton, New lower bound techniques for VLSI. Math. Syst. Theor. **17**, 47–70 (1984)
74. A. Mirzaian, Hamiltonian triangulations and circumscribing polygons of disjoint line segments. Comput. Geom. Theor. Appl. **2**(1), 15–30 (1992)
75. H. Nagamochi, T. Ibaraki, Augmenting edge-connectivity over the entire range in $O(nm)$ time. J. Algorithm. **30**, 253–301 (1999)

76. G. Narasimhan, M. Smid, *Geometric Spanner Networks* (Cambridge University Press, Cambridge, 2007)
77. V. Neumann-Lara, E. Rivera-Campo, J. Urrutia, A note on convex decompositions of point sets in the plane. *Graphs and Combinatorics* **20**(2), 223–231 (2004)
78. M. Newborn, W.O.J. Moser, Optimal crossing-free Hamiltonian circuit drawings of K_n. J. Combin. Theor. Ser. B **29**, 13–26 (1980)
79. W. Mulzer, G. Rote, Minimum-weight triangulation is NP-hard. J. ACM **55**(2), article 11 (2008)
80. J. O'Rourke, *Visibility*, ed. by J.E. Goodman, J. O'Rourke. Handbook of Discrete and Computational Geometry, 2nd edn. (CRC Press, Boca Raton, FL, 2004), pp. 643–665
81. J. O'Rourke, J. Rippel, Two segment classes with Hamiltonian visibility graphs. Comput. Geom. Theor. Appl. **4**, 209–218 (1994)
82. J. Pach, E. Rivera, Two segment classes with Hamiltonian visibility graphs. Comput. Geom. Theor. Appl. **10**, 121–124 (1998)
83. M. Patrignani, On extending a partial straight-line drawing. Int. J. Found. Comput. Sci. **17**(5), 1061–1069 (2006)
84. D. Rappaport, Computing simple circuits from a set of line segments is NP-complete. SIAM J. Comput. **18**(6), 1128–1139 (1989)
85. D. Rappaport, H. Imai, G.T. Toussaint, Computing simple circuits from a set of line segments. Discrete Comput. Geom. **5**(3), 289–304 (1990)
86. A. Rosenthal, A. Goldner, Smallest augmentations to biconnect a graph. SIAM J. Comput. **6**, 55–66 (1977)
87. G. Rote, F. Santos, I. Streinu, in *Pseudo-triangulations—A Survey*, ed. by J.E. Goodman, J. Pach. Surveys on Discrete and Computational Geometry: Twenty Years Later, vol 453 of Contemporary Mathematics (AMS, Providence, RI, 2008), pp. 343–410
88. T. Roughgarden, On the severity of Braess's paradox: Designing networks for selfish users is hard. J. Comput. Syst. Sci. **72**(5), 922–953 (2006)
89. T. Roughgarden, in *Routing Games*, ed. by N. Nisan et al. Algorithmic Game Theory, Chapter 18 (Cambridge University Press, Cambridge, 2007)
90. I. Rutter, A. Wolff, Augmenting the connectivity of planar and geometric graphs. Technical report, Universität Karlsruhe, 2008. Number 2008–3. http://digbib.ubka.uni-karlsruhe.de/volltexte/1000007814
91. J.R. Sack, J. Urrutia (eds.), *Handbook of Computational Geometry* (North Holland, Amsterdam, 2000)
92. T. Sakai, J. Urrutia, Convex decompositions of point sets in the plane, in *Proceedings of the 7th Japan Conference on Computational Geometry and Graphs*, JAIST, 2009
93. W. Schnyder, Embedding planar graphs on the grid, in *Proceedings of the 1st ACM-SIAM Symposium on Discrete Algorithms*, ACM Press. 1990, pp. 138–148
94. M. Sharir, A. Sheffer, Counting triangulations of planar point sets. Electron. J. Combinat. **18**, P70 (2011)
95. M. Sharir, A. Sheffer, E. Welzl, Counting plane graphs: perfect matchings, spanning cycles, and Kasteleyn's technique, *in Proceedings of the 28th ACM Symposium on Computational Geometry*, (ACM Press, 2012), pp. 189–198.
96. M. Sharir, E. Welzl, On the number of crossing-free matchings (cycles, and partitions). SIAM J. Comput. **36**(3), 695–720 (2006)
97. A. Sheffer, Numbers of plane graphs, http://www.cs.tau.ac.il/~sheffera/counting/PlaneGraphs.html
98. D.L. Souvaine, C.D. Tóth, A vertex-face assignment for plane graphs. Comput. Geom. Theor. Appl. **42**(5), 388–394 (2009)
99. I. Streinu, A combinatorial approach to planar non-colliding robot arm motion planning, in *Proceedings of the 41st Symposium on Foundations of Computer Science (FoCS)* (IEEE, Los Alamitos, CA, 2000), pp. 443–453
100. I. Streinu, Pseudo-triangulations, rigidity and motion planning. Discrete Comput. Geom. **34**, 587–635 (2005)

101. R. Tamassia (ed.), *Handbook of Graph Drawing and Visualization* CRC Press, 2013
102. E. Tardos, T. Wexler, in *Inefficiency of Equilibria in Network Formation Games*, ed. by N. Nisan et al. Algorithmic Game Theory, Chapter 19 (Cambridge University Press, Cambridge, 2007)
103. C.D. Tóth, Alternating paths along axis-parallel segments. Graphs Combinator. **22**(4), 527–543 (2006)
104. C.D. Tóth, Connectivity augmentation in planar straight line graphs. *Europ. J. of Combinatorics*, **33**(3), 408–425 (2012)
105. C.D. Tóth, P. Valtr, Augmenting the edge connectivity of planar straight line graphs to three, in *Proceedings of the 13th Spanish Meeting on Computational Geomentry*, Zaragoza, 2009
106. M. Urabe, M. Watanabe, On a counterexample to a conjecture of Mirzaian. Comput. Geom. Theor. Appl. **2**(1), 51–53 (1992)
107. J. Urrutia, in *Art Gallery and Illumination Problems*, ed. by J.R. Sack, J. Urrutia. Handbook on Computational Geometry (Elsevier, Amsterdam, 2000), pp. 973–1127
108. L. Végh, Augmenting undirected node-connectivity by one, in *Proceedings of the 42nd Symposium on Theory of Computing (STOC)* (ACM, New York, 2010), pp. 563–572
109. T. Watanabe, A. Nakamura, Edge-connectivity augmentation problems. J. Comput. Syst. Sci. **35**, 96–144 (1987)
110. E. Welzl, The number of triangulations on planar point sets, in *Proceedings of the 14th International Symposium (GD'06)*, LNCS, vol. 4372 (Springer, Berlin, 2007), pp. 1–4
111. C. Wulff-Nilsen, Computing the dilation of edge-augmented graphs in metric spaces. Comput. Geom. Theor. Appl. **43**, 68-72 (2010)
112. G. Xia, Improved upper bound on the stretch factor of Delaunay triangulations, in *Proceedings of the 27th Symposium on Computational Geomentry (SoCG)* (ACM, New York, 2011), pp. 264–273
113. G. Xia, L. Zhang, Improved lower bound for the stretch factor of Delaunay triangulations, in *Proceedings of the 20th Fall Workshop on Computational Geomentry* (Stony Brook, NY, 2010)

Discrete Geometry on Red and Blue Points in the Plane Lattice

Mikio Kano and Kazuhiro Suzuki

Abstract We consider some problems on red and blue points in the plane lattice. An L-line segment in the plane lattice consists of a vertical line segment and a horizontal line segment having a common endpoint. There are some results on geometric graphs on a set of red and blue points in the plane. We show that some similar results also hold for a set of red and blue points in the plane lattice using L-line segments instead of line segments. For example, we show that if n red points and n blue points are given in the plane lattice in general position, then there exists a noncrossing geometric perfect matching covering them, each of whose edges is an L-line segment and connects a red point and a blue point.

1 Introduction

We consider some problems on red points and blue points in the plane lattice $\mathbb{Z}^2$ motivated by some results in the plane $\mathbb{R}^2$, where $\mathbb{Z}$ and $\mathbb{R}$ denote the set of integers and the set of real numbers, respectively. For a point x in the plane, an L-shaped line consisting of a vertical ray and a horizontal ray emanating from x is called an *L-line* with *corner* x. A vertical line and a horizontal line passing through x are also considered special L-lines with corner x. So, for every point x, there are exactly six L-lines with corner x, and two of them are usual lines (see in Fig. 1).

M. Kano (✉)
Department of Computer and Information Sciences, Ibaraki University, Hitachi, Ibaraki 316-8511, Japan
e-mail: kano@mx.ibaraki.ac.jp; gorogoro.cis.ibaraki.ac.jp

K. Suzuki
Department of Information Science, Kochi University, Kochi 780-8520, Japan
e-mail: kazuhiro@tutetuti.jp

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_18,

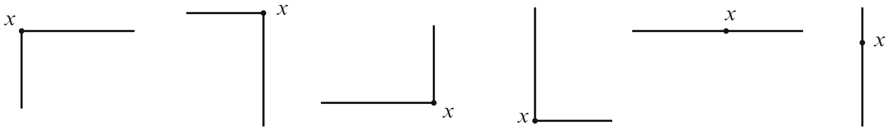


Fig. 1 *L*-lines with corner *x*

We regard *L*-lines as "lines" in the plane lattice, and consider some problems from this point of view. For two points in the plane lattice that are not on the same vertical or horizontal line, there are two *L*-lines passing through them [see (1) in Fig. 2]. On the other hand, for any two points in the plane, there exists exactly one line that passes through them. So there is a difference between lines in the plane and *L*-lines in the plane lattice. However, as we shall show, they have some nice common properties.

Remark. Let S be a set of points in the plane lattice. Usually, S is defined to be in general position if every vertical line or horizontal line contains at most one point of S. On the other hand, by using *L*-lines, we can define S to be in general position if no three points of S lie on the same *L*-line [see (2) in Fig. 2]. If S is in general position by means of the new definition, then the highest point and the lowest point of S may lie on the same vertical line. However, for any other point x of S, the vertical line passing through x does not pass through any point of $S-\{x\}$. Similarly, the rightmost and leftmost points of S may lie on the same horizontal line, but for any other point y of S, the horizontal line passing through y does pass through any point of $S-\{y\}$. Therefore, the above two definitions of general position are slightly different, but they require the same condition for most points in S, and thus the difference is small.

Hereafter, to avoid confusion and for simplicity, we say that S is *in general position* if every vertical line and horizontal line pass through at most one point of S. Namely, we use a standard definition of general position.[1]

2 Geometric Alternating Matchings

A set X of points in the plane is called *in general position* if no three points of X lie on the same line. It is well known that if n red points and n blue points are given in the plane in general position, then there exists a noncrossing geometric perfect

[1] In the plane, no three points lie on the same line if and only if every three points make a triangle. Similarly, in the plane lattice, every vertical line and horizontal line pass through at most one point if and only if every two points make a digon with two *L*-line segments.

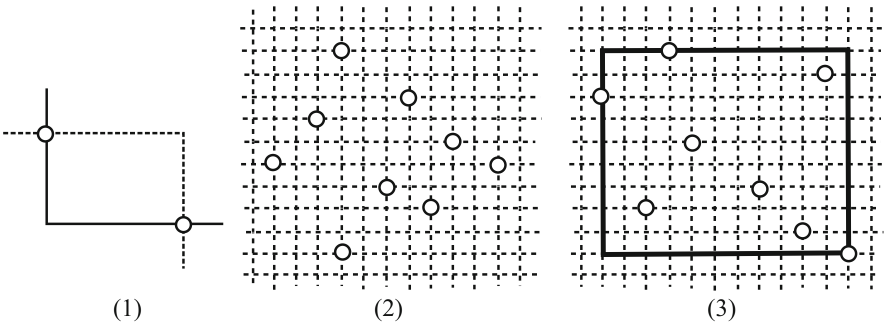
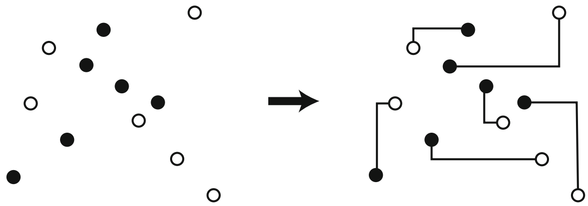


Fig. 2 (*1*) Two *L*-lines passing through two given points. (*2*) A set of points no three of which lie on the same *L*-line. (*3*) A rectangular hull of a set of points in the plane lattice in general position

Fig. 3 A noncrossing perfect matching with *L*-line segments joining *red* points and *blue* points



matching joining the red points and the blue points, where a geometric matching is a matching consisting of line segments. We start with a result on perfect matchings with L-line segments in the plane lattice.

Theorem 1. *Suppose that n red points and n blue points are given in the plane lattice in general position, where* $n \geq 1$ *is an integer. Then there exists a noncrossing perfect matching with L-line segments joining the red points and the blue points (Fig. 3).*

We need some new notation. In this chapter, only axis-parallel rectangles will be used, and so a *rectangle* always means such a rectangle. Thus, each edge of an rectangle is a vertical or horizontal line segment. For a set S of points in the plane lattice, the *rectangular hull* of S, denoted by $\mathrm{rect}(S)$, is the smallest closed rectangular enclosing S [(3) in Fig. 2)] In particular, every edge of $\mathrm{rect}(S)$ contains at least one point of S. For a set X, the cardinality of X is denoted by $|X|$ or $\#X$.

Proof of Theorem 2.2. We prove Theorem 1 by induction on n. If $n = 1$, then the theorem holds. So we assume $n \geq 2$. Let X be the set of points of S on the boundary of the rectangular hull $\mathrm{rect}(S)$. Then $2 \leq |X| \leq 4$. Suppose that X contains both a red point and a blue point. Then there exists an L-line segment L_1 that is on the boundary of $\mathrm{rect}(S)$ and joins a red point x to a blue point y of X. By applying the induction hypothesis to $S - \{x, y\}$, we obtain a noncrossing perfect matching with L-line segments joining the red points and the blue points of $S - \{x, y\}$. By adding L_1 to this matching, we can get the desired noncrossing perfect matching.

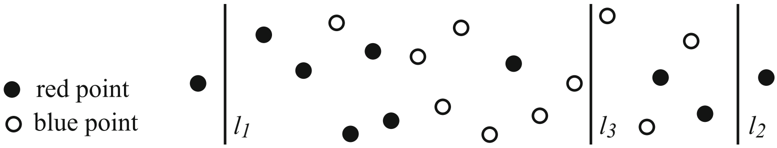


Fig. 4 Moving a *vertical line* ℓ from ℓ_1 to ℓ_2 to find ℓ_3 such that $f(\ell_3) = 0$

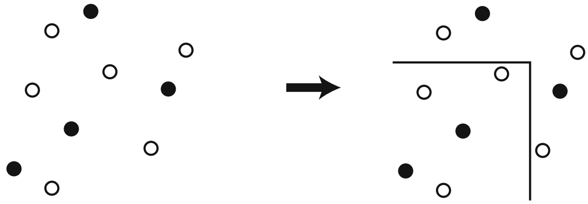


Fig. 5 An *L*-line that bisects both *red* points and *blue* points

Next, assume that all the points of X have the same color. By symmetry, we may assume that all the points of X are red. For every vertical line ℓ in the plane passing through no points of S, define a function $f(l)$ by

$$f(\ell) = \#\{\text{the red points of } S \text{ to the left of } \ell\}$$
$$-\#\{\text{the blue points of } S \text{ to the left of } \ell\}.$$

Then $f(\ell_1) = 1$ for a vertical line ℓ_1 immediately to the right of the left vertical edge of rect(S), and $f(\ell_2) = -1$ for a vertical line ℓ_2 immediately to the left of the right vertical edge of rect(S). Moreover, we continuously move a vertical line ℓ from ℓ_1 to ℓ_2. Then $f(\ell)$ changes by ± 1 when ℓ crosses a point of S. Hence, there exists a vertical line ℓ_3 such that $f(\ell_3) = 0$ and ℓ_3 passes through no point of S. By applying the induction hypothesis to the points of S to the left of ℓ_3 and those of S to the right of ℓ_3, respectively, we can obtain the desired noncrossing perfect matching (see Fig. 4). □

3 Balanced Subdivisions

We now turn our attention to another well-known theorem, the so-called ham-sandwich theorem, which says that if $2m$ red points and $2n$ blue points are given in the plane in general position, then there exists a line that bisects both red points and blue points. A similar result related to this theorem also holds by using *L*-lines, as in the following theorem.

Theorem 2 ([4,7]). *Let $m \geq 1$ and $n \geq 1$ be integers. If $2m$ red points and $2n$ blue points are given in the plane lattice in general position, then there exists an L-line that bisects both red points and blue points (Fig. 5).*

The above Theorem 2 was generalized as follows.

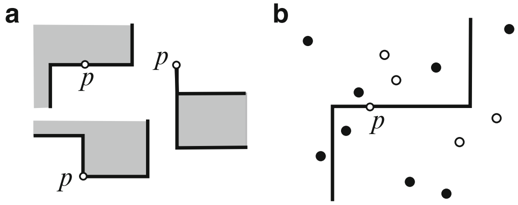


Fig. 6 (*1*) Three types of two *L*-rays emanating from *p*, which subdivide the plane into two regions. (*2*) Two *L*-rays emanating from *p* that bisect both *red* points and *blue* points

Theorem 3 (Bereg [2]). *Suppose that km red points and kn blue points are given in the plane lattice in general position, where* $m \geq 1$*,* $n \geq 1$*, and* $k \geq 2$ *are integers. Then there exists a subdivision of the plane into k regions with at most* $k-1$ *horizontal line segments and at most* $k-1$ *vertical line segments such that every region contains precisely n red points and m blue points.*

Bárány and Matoušek obtained the following theorem about another bisection.

Theorem 4 (Bárány and Matoušek [1]). *Suppose that* $2m$ *red and* $2n$ *blue points are given in the plane in general position, where* $m \geq 1$ *and* $n \geq 1$ *are integers. Let p be a point in the plane such that the red points, the blue points, and p are in general position. Then there exist two rays emanating from p that bisect both red points and blue points.*

A ray in the plane is a half-line emanating from a point. Similarly, an *L-ray* in the plane is defined to be a half *L*-line emanating from a point, and so an *L*-ray has a corner and an endpoint (see Fig. 6). We show that a similar result holds in the plane lattice using *L*-rays.

Theorem 5. *Suppose that* $2m$ *red points and* $2n$ *blue points are given in the plane lattice in general position, where* $m \geq 1$ *and* $n \geq 1$ *are integers. Let p be a point in the plane each of whose coordinates is not an integer. Then there exist two L-rays emanating from p that bisect both red points and blue points (see Fig. 6).*

Proof. First, take a big rectangle $\mathcal{R}$ that contains all the red points, the blue points, and p. We subdivide $\mathcal{R}$ into four regions by the horizontal line and the vertical line passing through p. Then for each given red or blue point x contained in the lower right region, we assign a new point y with the same color as x on the right edge of $\mathcal{R}$ such that x and y lie on the same horizontal line (see Fig. 7). For every given point x' in other regions, we assign a point y' with the same color as x' on the boundary of $\mathcal{R}$, as shown in Fig. 7.

Since given points are in general position, the assignment defined above is a bijection. It is easy to see that the boundary of $\mathcal{R}$ can be divided into two continuous parts so that each part contains precisely m red points and n blue points (see two bold marks on the boundary of $\mathcal{R}$ in Fig. 7). Notice that the reader is referred to [3], for example, about the proof of this fact. Then we can obtain the desired two L-rays, which emanate from the point p and pass through the partitioning marks on the boundary. □

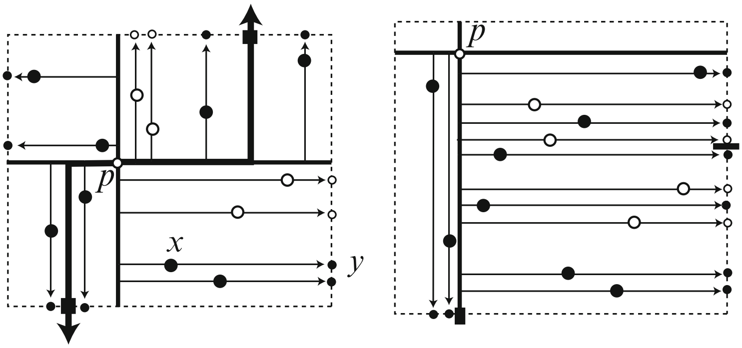


Fig. 7 The given *red* and *blue* points are assigned to *red* and *blue* points, respectively, on the boundary of $\mathcal{R}$

4 Geometric Spanning Trees

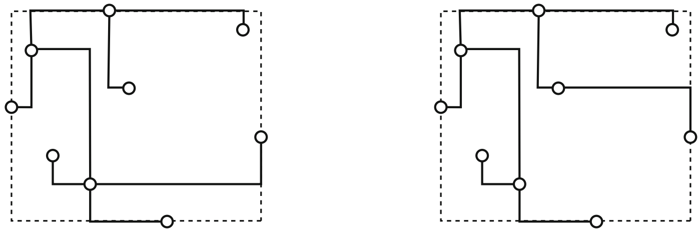
For a set X of points in the plane, we can draw noncrossing geometric spanning trees on X, each of whose edges is a line segment joining two points of X. In this chapter, we call such a spanning tree an *X-tree*. Given a set R of red points and a set B of blue points in the plane in general position, the minimum number of crossings of an R-tree and a B-tree is determined in the next theorem, where $\mathrm{conv}(X)$ denotes the convex hull of X.

Theorem 6 (Tokunaga [6]). *Let R and B be two disjoint sets of red points and blue points such that $R \cup B$ is in general position. Let $\tau(R,B)$ denote the number of edges xy of the boundary of $\mathrm{conv}(R \cup B)$ such that one of $\{x,y\}$ is red and the other is blue. Then $\tau(R,B)$ is even, and the minimum number of crossings in $T_{\mathrm{R}} \cup T_{\mathrm{B}}$ among all pairs $\{R$-tree T_{R}, B-tree $T_{\mathrm{B}}\}$ is equal to*

$$\max\left\{\frac{\tau(R,B)-2}{2},\ 0\right\}.$$

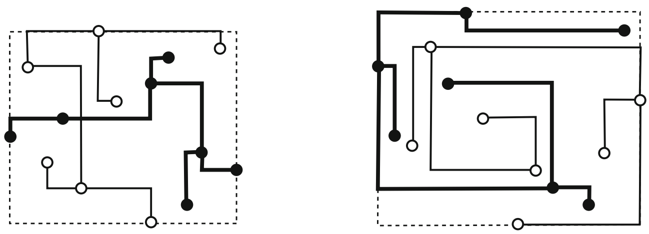
In particular, we can draw an R-tree and a B-tree without crossings if and only if $\tau(R,B) \le 2$.

We consider a similar problem on the plane lattice and prove a similar result as shown in the following Theorem 7. For a set X of points in the plane lattice in general position, we can draw noncrossing spanning trees on X each of whose edges is an L-line segment connecting two points of X, which is called a *spanning tree on X with L-line segments*, or simply an *X-tree with L-line segments*. Hereafter, we consider only noncrossing spanning trees with L-line segments, and so X-tree means X-tree with L-line segments (see Fig. 8). For a tree T and a vertex $v \in V(T)$, the degree of v in T is denoted by $\deg_T(v)$. The maximum degree of T is denoted by $\Delta(T)$.



(1) X-tree (2) X-tree of degree at most 3

Fig. 8 Examples of X-trees in the plane lattice



(1) $\tau*(R,B)=4$ (2) $\tau*(R,B)=2$

Fig. 9 Two spanning trees, R-tree and B-tree, with a minimum number of crossings

Theorem 7. *Let R and B be two disjoint sets of red points and blue points in the plane lattice such that $R \cup B$ is in general position. Let $\tau^*(R,B)$ denote the number of L-line segments xy on the boundary of* $\text{rect}(R \cup B)$ *such that one of $\{x,y\}$ is red and the other is blue. Then $\tau^*(R,B)$ is* 0*,* 2*, or* 4*, and the minimum number of crossings in $T_R \cup T_B$ among all pairs $\{R$-tree T_R, B-tree $T_B\}$ is equal to 1 if $\tau^*(R,B) = 4$. Moreover, if $\tau^*(R,B) \leq 2$, then we can draw an R-tree T_R and a B-tree T_B without crossings such that $\Delta(T_R) \leq 3$ and $\Delta(T_B) \leq 3$ (see Fig. 9).*

Note that if $\tau^*(R,B) = 4$, then any T_R and T_B cross at least once. In fact, let $x_r \in R$, $y_r \in R$, $x_b \in B$, and $y_b \in B$ be the left, right, top, and bottom points in $\text{rect}(R \cup B)$, respectively. The path in T_R starting from x_r to y_r and the path in T_B starting from x_b to y_b cross at least once.

In order to prove Theorem 7, we need some definitions and a lemma. An *orthogonal spiral polygon* is a polygon whose boundary consists of two chains of edges, which are called the *outer chain* and *inner chain*, respectively. Every internal angle of the outer chain is $\pi/2$, and every internal angle of the inner chain is $3\pi/2$ [see (1) in Fig. 10]. Notice that we allow an edge of the inner chain to be included in an edge of the outer chain; namely, some part of the polygon may consist of only edges (no inner points) and be flattened.

Lemma 8. *Let $\mathcal{P}$ be an orthogonal spiral polygon in the plane lattice, and let S be a set of points in the plane lattice in general position contained in $\mathcal{P}$. Assume that every edge of the outer chain of $\mathcal{P}$ contains exactly one point of S and that every*

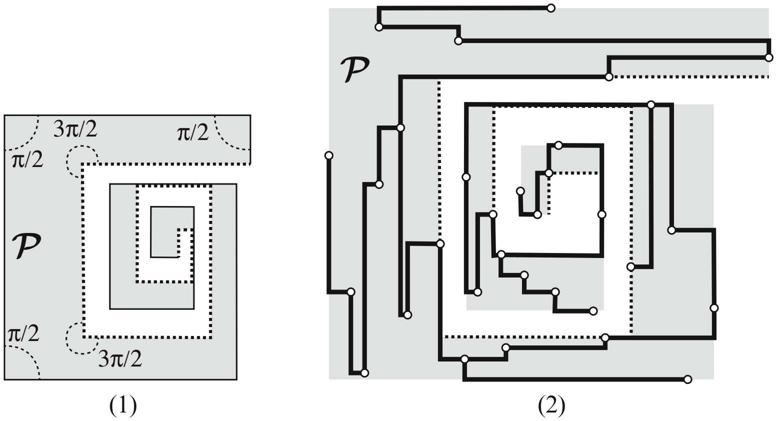


Fig. 10 (*1*) An orthogonal spiral polygon $\mathcal{P}$ with an outer chain and an inner chain consisting of *bold* and *broken* edges, respectively. (*2*) An orthogonal spiral polygon $\mathcal{P}$ with a point set S and an S-tree T such that $\Delta(T) \leq 3$

edge of the inner chain contains exactly one point of S or is included in another edge of the outer chain. Then there exists an S-tree T such that (i) $\Delta(T) \leq 3$ and (ii) T is included in $\mathcal{P}$ [see (2) in Fig. 10].

Proof. An edge of the inner chain of $\mathcal{P}$ included in another edge of the outer chain is called a *flattened rectangle*. Note that a flattened rectangle may have one point of S by our assumption for $\mathcal{P}$. In (2) of Fig. 10, we can find two flattened rectangles such that one of them has one point of S and another one has no points of S. We may assume that no point of S is at a corner of the inner chain of $\mathcal{P}$ if the corner is not contained in an edge of the outer chain. Otherwise, we may move one of the edges incident to the corner so that the resulting orthogonal spiral polygon $\mathcal{P}'$ is included in $\mathcal{P}$. In other words, we consider a minimal orthogonal spiral polygon included in $\mathcal{P}$ on S.

In preparation for our construction of an S-tree, we decompose the orthogonal spiral polygon $\mathcal{P}$ into closed rectangles as shown in Fig. 11. If $\mathcal{P}$ has no flattened rectangles, then we decompose $\mathcal{P}$ as shown (1) in Fig. 11, where X, Y, and Z denote closed rectangles. The top edge of Y is included in the bottom edge of X, and the left edges of X and Y form an edge of the outer chain of $\mathcal{P}$. Two consecutive rectangles Y and Z have the same properties, and so on.

If $\mathcal{P}$ has some flattened rectangles, then we decompose $\mathcal{P}$ as shown in (2) in Fig. 11, where each X_i denotes each closed rectangle. Some rectangles (X_3, X_5, X_6, X_8 in Fig. 11) are flattened. Remove these flattened rectangle from $\mathcal{P}$. Then the remaining parts of $\mathcal{P}$ consist of some orthogonal spiral polygons, so we decompose each of these as shown in (1) in Fig. 11.

We denote these decomposed rectangles by X_1, X_2, ..., and X_k in spiral order.

Claim 1. No three flattened rectangles without points of S are consecutive in $\mathcal{P}$.

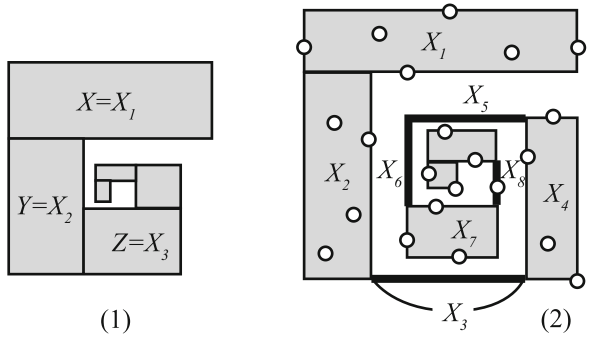


Fig. 11 (*1*) A decomposition of an orthogonal spiral polygon without flattened rectangles. (*2*) A decomposition of an orthogonal spiral polygon with flattened rectangles, where X_3, X_5, X_6, and X_8 are flattened rectangles and only X_8 contains one point of S

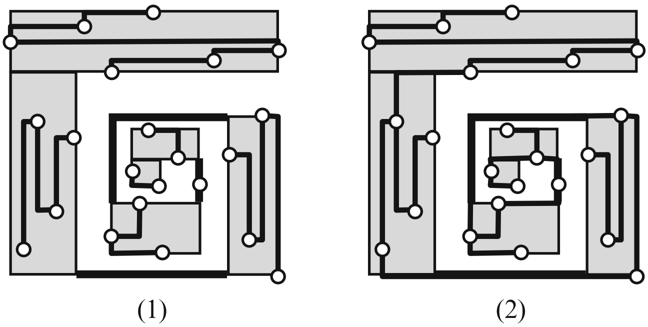


Fig. 12 An outline of our construction of an S-tree

Proof. Otherwise, some edge of the outer chain has no points of S, which contradicts our assumption. □

Figure 12 shows an outline of our construction of an S-tree. First, construct a path in each nonflattened rectangle X_i from "outer" point to "inner" point as shown (1). Then we connect those paths as shown in (2).

We shall show more precisely how to do it. Each nonflattened rectangle X_i consists of four edges. Exactly one of these edges includes an edge of the inner chain of $\mathcal{P}$. We regard this edge as the *bottom* edge of the rectangle. Then, in accordance with the bottom edge, we can define a *top*, *right*, and *left* edge of the rectangle. Let p_i, q_i, r_i, and t_i be the most top, bottom, right, and left points of $X_i \cap S$, respectively. We construct a path P_i with L-line segments that satisfies the following three properties: (1) P_i starts at p_i and ends at q_i; (2) P_i passes through all the points in $X_i \cap S$ from top to bottom; (3) each L-line segment xy such that x is upper than y starts at x to the right or left and ends at y from above (see Fig. 13).

For each flattened rectangle X_i with one point $x_i \in S$, let $p_i = q_i = r_i = t_i = x_i$ and $P_i = \{x_i\}$. For each flattened rectangle X_i without points of S, let a_i be the center point on X_i, $p_i = q_i = r_i = t_i = a_i$, and $P_i = \emptyset$. We will use the points a_i as dummy points of P_i.

Next, we connect each two paths P_i and P_{i+1} as follows.

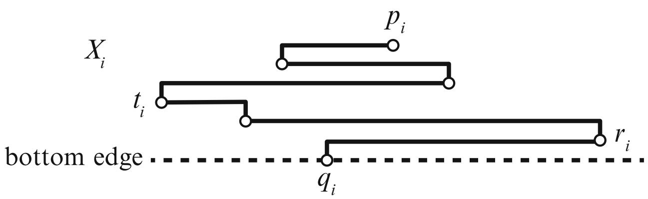


Fig. 13 A path P_i on X_i

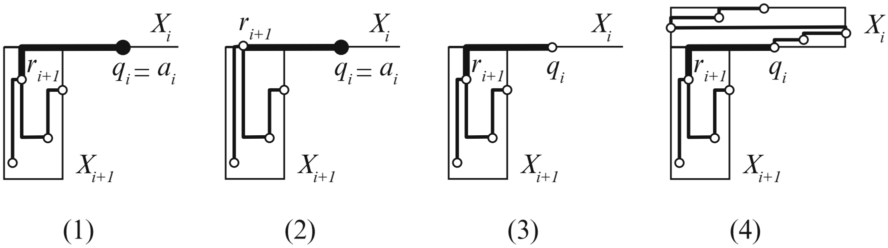


Fig. 14 How to connect each two paths P_i and P_{i+1} if $P_{i+1} \neq \emptyset$. The *black points* are dummy points

Case 1. $P_{i+1} \neq \emptyset$.

In this case, we can always connect the bottom point q_i and the right point r_{i+1} with an L-line segment or a line segment without crossings as follows. If $P_i = \emptyset$ or $|P_i| = 1$, then X_i is a flattened rectangle. Thus, by the definition of decomposition of $\mathcal{P}$, the bottom point q_i is on an edge of the inner chain of $\mathcal{P}$. Therefore, we can connect q_i and r_{i+1} without crossings, as shown in Fig. 14(1)–(3).

If $|P_i| \geq 2$, then, similarly, the bottom point q_i of P_i is on an edge of the inner chain of $\mathcal{P}$, since otherwise some edge of the inner chain has no points of S and is not included in an edge of the outer chain, which contradicts our assumption. Hence, in X_i, a horizontal line segment from q_i to the left can be added to the path P_i for connecting P_i and P_{i+1}, and this segment does not overlap with another segment from X_{i-1}, namely, from the right. Therefore, we can connect q_i and r_{i+1} without crossings, as shown in Fig. 14(4). Note that if $|P_{i+1}| \geq 2$ and $q_i = r_{i+1}$, that is, q_i is on a corner of the inner chain, then the degree $\deg_T(q_i)$ may be 4. However, we assumed that no point of S is at a corner of the inner chain of $\mathcal{P}$ if the corner is not contained in an edge of the outer chain.

Case 2. $P_{i+1} = \emptyset$.

In this case, we can always connect the left point t_i and the dummy point a_{i+1} with an L-line segment or a line segment without crossings, as shown in Fig. 15.

Consequently, by ignoring the dummy points a_i, we get a tree T on S without crossings such that (1) $\Delta(T) \leq 3$ and (2) T is included in $\mathcal{P}$. We shall show that T is an S-tree; namely, every edge of T is an L-line segment.

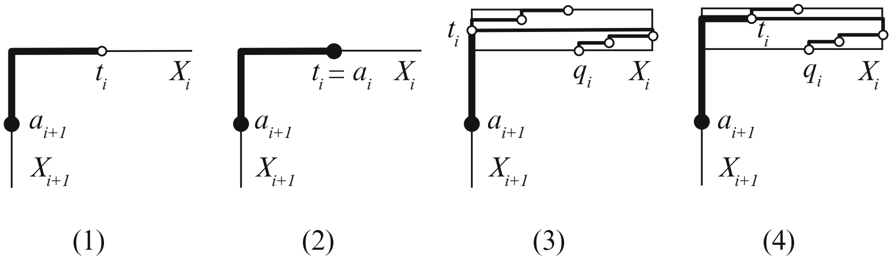


Fig. 15 How to connect each two paths P_i and P_{i+1} if $P_{i+1} = \emptyset$

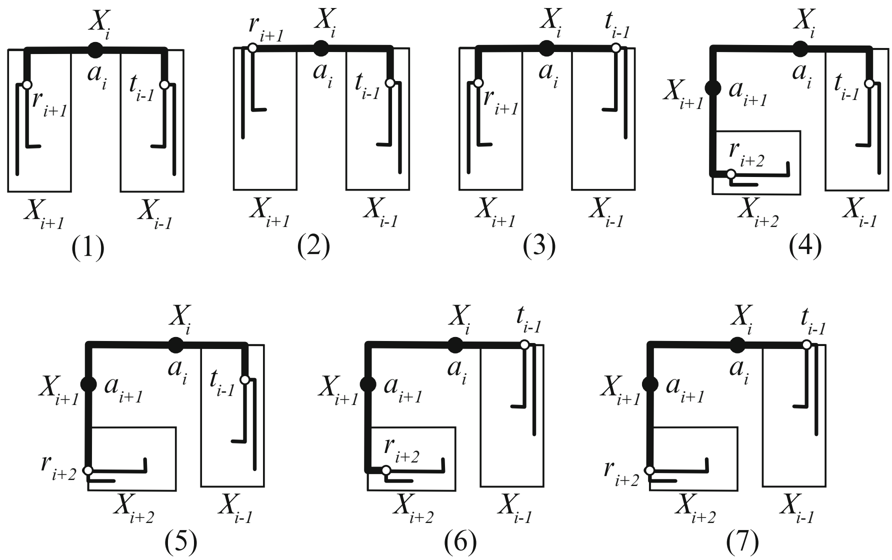


Fig. 16 Seven cases of edges of T through dummy points

The edges of T not through dummy points are L-line segments. Thus, we consider edges of T through dummy points. By Claim 1, there are just seven cases, as shown in Fig. 16, where the bold edges are edges of T through one or two dummy points.

The cases (1), (4)–(6) in Fig. 16 contradict our assumption that every edge of the souter chain of $\mathcal{P}$ contains exactly one point of S. In the other cases, each bold edge is an L-line segment. Therefore, the tree T is a desired S-tree. Consequently, the lemma is proved. □

Proof of Theorem 4.7. We first prove the theorem in the case where the top point and the leftmost point in rect$(R \cup B)$ are red and the bottom point and the rightmost point in rect$(R \cup B)$ are blue, in particular, $\tau^*(R,B) = 2$.

We take some rectangles containing only red points or blue points and obtain two disjoint orthogonal spiral polygons that contain all the red points and all the blue

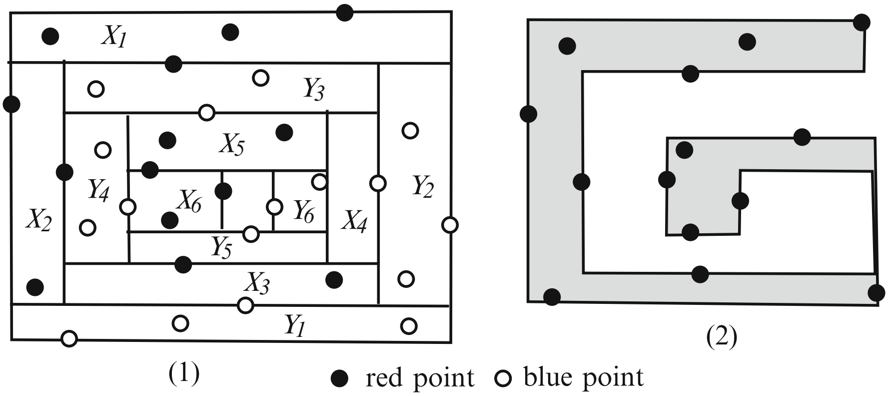


Fig. 17 (*1*) Rectangles $X_1, X_2, \ldots, X_6$ and $Y_1, Y_2, \ldots, Y_6$. (*2*) An orthogonal spiral polygon containing all the *red* points, which is included in $X_1 \cup \ldots \cup X_6$

points, respectively. First, take the largest rectangle X_1 that contains no blue points, whose top edge is the top edge of rect$(R \cup B)$, and whose bottom edge contains a red point (see Fig. 17). Next, take the largest rectangle Y_1 that contains no red points, whose bottom edge is the bottom edge of rect$(R \cup B)$, and whose top edge contains a blue point. Then remove open region $X_1 \cup Y_1$ together with the red and blue points in $X_1 \cup Y_1$ from rect$(R \cup B)$, and denote the resulting rectangle by Rect$_2$, whose red point set is R_2 and blue point set is B_2.

Hereafter we assume that rectangle X_i contains no blue points and rectangle Y_i contains no red points. Take the largest rectangle X_2 whose left edge is the left edge of Rect$_2$ and whose right edge contains a red point if any. Namely, if the leftmost point of Rect$_2$ is blue [i.e., this case may occur if the leftmost point in rect$(R \cup B)$ lies on the left edge of X_1], then X_2 consists of only one edge and contains no inner points and no red points. Similarly, take the largest rectangle Y_2 whose right edge is the right edge of Rect$_2$ and whose left edge contains a blue point, if any. So it may occur that Y_2 consists of one edge and contains no blue points. Then remove open region $X_2 \cup Y_2$ and the red and blue points in $X_2 \cup Y_2$ from Rect$_2$, and denote the resulting rectangle by Rect$_3$, whose point set is $R_3 \cup B_3$. Note that if X_2 contains no red points, then Rect$_3$ is obtained from Rect$_2$ only by removing Y_2. Moreover, if X_2 contains no red points and Y_2 contains no red blue points, then Rect$_3$ is equal to Rect$_2$, but we next take the largest rectangle X_3 whose bottom edge is the bottom edge of Rect$_3$ and whose top edge contains a red point, and X_3 contains at least one red point.

We repeat the same procedure until rect$(R_k \cup B_k)$ contains neither red points nor blue points [see (1) of Fig. 17]. Then $X_1 \cup X_2 \cup \ldots \cup X_k$ is an orthogonal spiral polygon containing all the red points. If an edge does not contain a red point, we move the edge to inside until it contains a red point or is included in another edge. By repeating this procedure, we can obtain the desired orthogonal spiral polygon, which contains all the points of R and each of whose edges either contains one red

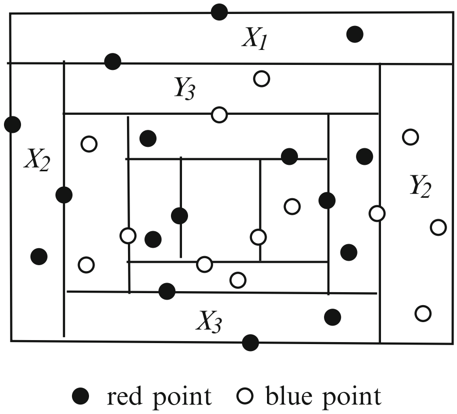


Fig. 18 The *top* point, the *leftmost* point, and the *bottom* point in rect$(R \cup B)$ are *red* and the *rightmost* point in rect$(R \cup B)$ is *blue*

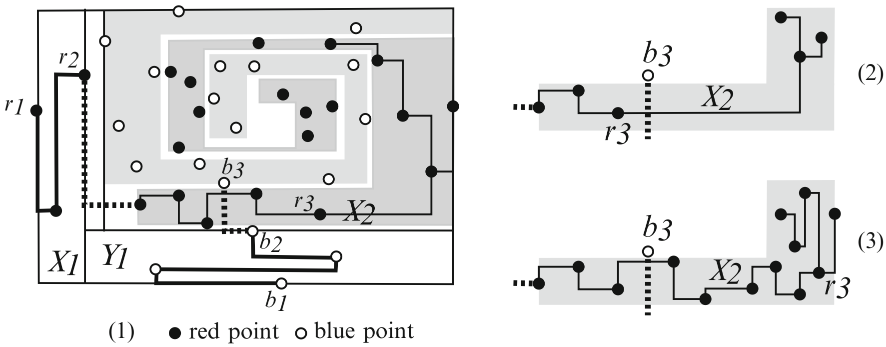


Fig. 19 An outline of our construction of an R-tree and a B-tree with exactly one crossing. Three cases on the place of r_3

point or is included in another edge. By Lemma 8, we can obtain a spanning tree with L-line segments on R with maximum degree at most 3. Similarly, we can obtain a blue spanning tree with maximum degree at most 3, and it is clear that these two spanning trees do not cross.

We next consider the case where $\tau^*(R,B) = 2$ and the top point, the leftmost point, and the bottom point in rect$(R \cup B)$ are red and the rightmost point in rect$(R \cup B)$ is blue (see Fig. 18).

We first take a rectangle X_1 as above. Then rect$(R \cup B) - X_1$ satisfies the condition of the case discussed above, and so we can apply the procedure to take $X_2, Y_2, X_3, Y_3, \dots, X_k, Y_k$. Then we can obtain two disjoint orthogonal spiral polygons from $X_1 \cup \cdots \cup X_k$ and $Y_2 \cup \cdots \cup Y_k$, respectively, and thus we can get the desired two spanning trees. In the other cases of $\tau^*(R,B) \leq 2$, we can similarly obtain the desired two spanning trees.

We finally consider the case where $\tau^*(R,B) = 4$, the rightmost point and the leftmost point of rect$(R \cup B)$ are red, and the top point and the bottom point of rect$(R \cup B)$ are blue (see Fig. 19).

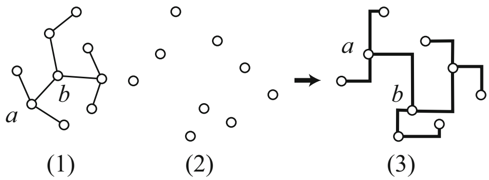


Fig. 20 (*1*) A tree T with maximum degree 3. (*2*) A set P of $|T|$ points in the plane lattice in general position. (*3*) T is drawn on P with L-line segments without crossing

First take the largest rectangle X_1 whose left edge is the left edge of rect$(R \cup B)$ and whose right edge contains a red point. Let r_1 and r_2 be the leftmost and the rightmost red points in X_1, respectively. We construct a path starting at r_1 ending at r_2 from left to right, as shown in (1) of Fig. 19. Then remove X_1 and their points from rect$(R \cup B)$, and denote the resulting rectangle by Rect$_1$, whose point set is $R_1 \cup B_1$.

Next take the largest rectangle Y_1 whose bottom edge is the bottom edge of Rect$_1$ and whose top edge contains a blue point. Let b_1 and b_2 be the bottom and top blue points in Y_1, respectively. We construct a path starting at b_1 from bottom to top as shown in (1) of Fig. 19. Then remove Y_1 and their points from Rect$_1$, and denote the resulting rectangle by Rect$_2$ with point set $R_2 \cup B_2$.

We take the largest rectangle X_2 whose bottom edge is the bottom edge of Rect$_2$ and whose top edge contains a red point. Let b_3 be the bottom point in B_2. We connect b_2 and b_3 by an L-line segment such that a horizontal line segment starts at b_2.

Let r_3 be the rightmost red point in X_2. Remove $X_2 \setminus \{r_3\}$ and their points from Rect$_2$, and denote the resulting rectangle by Rect$_3$ with point set $R_3 \cup B_3$. Then $\tau^*(R_3, B_3) = 2$, and so there exist a red R_3-tree T'_{R} and a blue B_3-tree T'_{B}. When we construct these two trees, we first take the largest rectangle that contains r_3.

By the construction method for an S-tree in the proof of Lemma 8, we can add a horizontal line segment to the left-hand side of r_3. Then, we construct a path starting at r_2 ending at r_3 from left to right, as shown in (1) of Fig. 19. Wherever r_3 is, exactly one L-line segment crosses the L-line segment b_2b_3 (see Fig. 19).

Consequently, we obtain the desired red spanning tree and blue spanning tree with exactly one crossing, and the proof is complete. □.

The degree of points of our R-tree and B-tree in the proof of Theorem 7 is at most 3 except b_3. We now give a problem concerning Theorem 7.

Problem 9. In Theorem 7, even if $\tau^*(R, B) = 4$, then is it possible to require that $\Delta(T_{\mathrm{R}}) \leq 3$ and $\Delta(T_{\mathrm{B}}) \leq 3$? Moreover, is it possible to replace a spanning tree with maximum degree 3 by a Hamilton path (i.e., a spanning tree with maximum degree 2)?

We conclude this chapter with the following conjecture. Let T be a tree and let P be a set of $|T|$ points in the plane lattice in general position, where $|T|$ denotes the order of T. If T can be drawn on P without crossing such that each edge of T is an L-line segment connecting two points of P, then we say that T can be drawn on P with L-line segments (see Fig. 20).

Conjecture 10. Let T be a tree with maximum degree 3, and let P be a set of $|T|$ points in the plane lattice in general position. Then T can be drawn on P with L-line segments without crossing.

A partial solution to this conjecture is given in [5]; namely, it is proved that if a tree T with maximum degree 3 has the property that all the vertices of degree 3 are contained in a path of T, then the conjecture holds.

References

1. I. Bárány, J. Matoušek, Simultaneous partitions of measures by k-fans. Discrete Comput. Geom. **25**, 317–334 (2001)
2. S. Bereg, Orthogonal equipartitions. Comput. Geom. **42**, 305–314 (2009)
3. A. Kaneko, M. Kano, A balanced interval of two sets of points on a line. *Combinatorial Geometry and Graph Theory*. LNCS, vol. 3330 (Springer, Berlin, 2005), pp. 108–112
4. M. Uno, T. Kawano, M. Kano, Bisections of two sets of points in the plane lattice, IEICE Transactions on Fundamentals vol. E92-A, 502–507 (2009)
5. M. Kano, K. Suzuki, Geometric Graphs in the Plane Lattice, XIV Spanish meeting on computational geometry, LNCS, Springer-Verlag, Berlin,7579, 204–281 (2012)
6. S. Tokunaga, Intersection number of two connected geometric graphs. Inform. Process. Lett. **59**, 331–333 (1996)
7. M. Kano, K. Suzuki, M. Uno, T. Kawano, M. Kano, Bisections of two sets of points in the plane lattice. IEICE Trans. **340**, Fundament. E92-A, 502507 (2009)

Ramsey-Type Problems for Geometric Graphs

Gyula Károlyi

Abstract We survey some results and collect a set of open problems related to graph Ramsey theory with geometric constraints.

1 Introduction

A *geometric graph* is a graph drawn in the plane so that every vertex corresponds to a point, and every edge is a closed straight-line segment connecting two vertices but not passing through a third. The $\binom{n}{2}$ segments determined by n points in the plane, no three of which are collinear, form a *complete* geometric graph with n vertices. A geometric graph is *convex* if its vertices correspond to those of a convex polygon. Further, we say that a subgraph of a geometric graph is *noncrossing* if no two of its edges have an interior point in common.

A systematic study of Ramsey-type problems for geometric graphs was initiated in [22, 23]. These questions were further investigated in [4, 9, 17, 20, 21, 24, 25]. The aim of this writing is to give a noncomprehensive survey of these developments, with an emphasis on open problems. We also prove a new result concerning noncrossing monochromatic matchings in multicolored geometric graphs and address the problem of Ramsey multiplicity.

A typical problem in Ramsey theory is the following. Given a finite sequence $G_1, G_2, \ldots, G_t$ of simple graphs, determine the smallest integer r, denoted by $R(G_1, G_2, \ldots, G_t)$, with the property that whenever the *edges* of a complete graph on at least r vertices are partitioned into t color classes, there is an integer $1 \leq i \leq t$ such that the ith color class contains a subgraph isomorphic to G_i. Such a subgraph is referred to as a monochromatic subgraph in the ith color. In the special case,

G. Károlyi (✉)
Institute of Mathematics, Eötvös University Pázmány P. sétány 1/c, Budapest, H–1117 Hungary
e-mail: karolyi@cs.elte.hu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_19,

when each $G_i = K_{k_i}$ is a complete graph on k_i vertices, we simply write $R(k_1,k_2,\dots,k_t)$ for $R(G_1,G_2,\dots,G_t)$. In general, if G_i has k_i vertices, then the existence of $R(G_1,G_2,\dots,G_t)$ follows directly from that of $R(k_1,k_2,\dots,k_t)$, first observed by Ramsey [29]. We may arrive at a somewhat broader concept of $R(\mathcal{G}_1,\mathcal{G}_2,\dots,\mathcal{G}_t)$ by insisting that, for some i, the ith color class contains a copy of a graph belonging to a prescribed family $\mathcal{G}_i$.

For a sequence of graphs $G_1,G_2,\dots,G_t$, the *geometric* Ramsey number $R_g(G_1,G_2,\dots,G_t)$ is defined as the smallest integer r with the property that whenever the edges of a complete geometric graph on at least r vertices are partitioned into t color classes, the ith color class contains a noncrossing copy of G_i, for some $1 \le i \le t$. The number $R_c(G_1,G_2,\dots,G_t)$ denotes the corresponding number if we restrict our attention to *convex* geometric graphs only. These numbers exist if and only if each graph G_i is *outerplanar*, that is, can be obtained as a subgraph of a triangulated cycle (convex n-gon triangulated by noncrossing diagonals). The necessity of the condition is obvious, whereas the "if part" is implied by the following result of Gritzmann et al. [15].

Theorem A. *Let P be an arbitrary set of n points in the plane in general position. For any outerplanar graph H on n vertices, there is a straight-line embedding f of H into the plane such that the vertex set of $f(H)$ is P and no two edges of $f(H)$ cross each other.*

Corollary B. $R(G_1,\dots,G_t) \le R_c(G_1,\dots,G_t) \le R_g(G_1,\dots,G_t) \le R(k_1,\dots,k_t)$ *holds for arbitrary outerplanar graphs $G_1,\dots,G_t$ with $k_1,\dots,k_t$ vertices, respectively.*

Most known results concern the diagonal bi-colored case, that is, when $t = 2$ and $G_1 = G_2$. For simplicity, write $R(G)$, $R_g(G)$, and $R_c(G)$ for $R(G,G)$, $R_g(G,G)$, and $R_c(G,G)$, respectively. Due to the inequality $R_c(G) \le R_g(G)$ and the cyclically ordered structure of convex complete geometric graphs, it is generally easier to obtain/prove upper bounds for R_c than to R_g. On the other hand, the largest number of crossing edges in a complete geometric graph occurs when the vertices are in convex position, suggesting that R_g should not be much larger than R_c.

Problem 1. Is it true that $R_g(G) = R_c(G)$ holds for every outerplanar graph G?

Most likely the answer is negative, but it seems difficult to find a counterexample. It would also be interesting to see a nontrivial upper bound.

2 Spanning Trees

For any finite graph G, either G or its complement $\overline{G}$ is connected. That is, either G or $\overline{G}$ contains a spanning tree. This observation extends to a geometric setting as follows; see [22]. In our terminology it means that for any integer $n \ge 2$, we have $R_g(\mathcal{T}_n) = n$, where $\mathcal{T}_n$ denotes the family of all trees with n vertices.

Theorem C. *If the edges of a finite complete geometric graph are colored by two colors, there exists a noncrossing spanning tree, whose edges are all the same color.*

For *convex* geometric graphs, this follows by simple induction on the number of vertices [2]. If all edges along the boundary of the convex hull are of the same color, then there is a monochromatic noncrossing spanning path. Otherwise, let ab and bc be two edges of the convex hull having a different color; omit the vertex b, and apply the induction hypothesis. The second part of the argument works also in the nonconvex case. The tricky part is when all the boundary edges are of the same color, say red. In that case, one can use a sweeping line, which either splits the graph into two parts sharing a vertex such that each part admits a noncrossing spanning tree of the same color, or splits the graph into two parts without a common vertex such that both parts admit red noncrossing spanning trees that can be merged along a boundary edge.

One cannot in general expect a monochromatic noncrossing spanning path (see below), but probably Theorem C can still be strengthened. A *caterpillar* is a tree obtained from a path and a set of isolated vertices, connecting each isolated vertex to the path with a new edge. It is known that any finite graph or its complement contains a spanning caterpillar, in fact, even a spanning broom (a path with additional edges attached to one of its endpoints) [6, 18]. Following up the investigations in [27], Micha Perles suggested that the following may be true, at least for convex geometric graphs.

Problem 2. The edges of a finite complete geometric graph are colored by two colors. Is it always true that there exists a noncrossing monochromatic spanning caterpillar?

Regarding this problem, Keszegh (personal communication, 2011) proved the following weaker version: If the edges of a finite complete geometric graph are colored red and blue, and the red subgraph is not connected, then the blue subgraph contains a noncrossing spanning caterpillar. He also noticed that there is not always a noncrossing monochromatic spanning broom.

A related problem concerning 2-colorings of complete *bipartite* graphs was suggested by Gyárfás [18]. It is not true that in any 2-coloring of a complete bipartite graph on n vertices, there would be a monochromatic spanning tree. On the other hand, a monochromatic tree with at least $n/2$ vertices always exists [16]. A *simple* geometric bipartite graph $K_{n,n}$ is obtained by placing $2n$ points along a circle for vertices, and connecting each of n consecutive vertices with each of the remaining ones with a straight-line segment.

Problem 3. Is it true that in any 2-coloring of a simple geometric $K_{n,n}$, there exists a noncrossing monochromatic subtree with n vertices?

Note that such a tree would necessarily be a caterpillar. As the dominant color class contains at least $n^2/2$ edges, from the following unpublished result of Perles, it is immediate that for any fixed caterpillar tree T with at most $n/2$ vertices, there is a noncrossing monochromatic subtree isomorphic to T.

Theorem D. *If a convex geometric graph of* $n \geq v$ *vertices has more than* $\lfloor (v-2)n/2 \rfloor$ *edges, then it contains a noncrossing copy of any caterpillar with* v *vertices.*

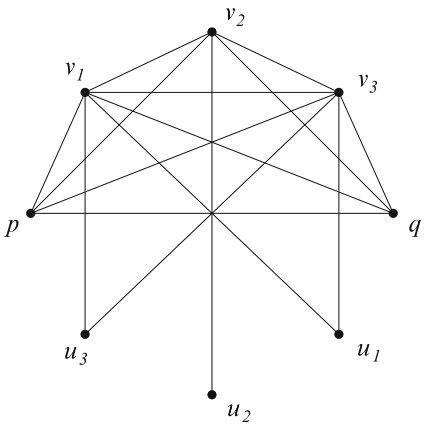
This sharp Turán-type result was first mentioned in [28]; a proof is available in [5]. More problems for noncrossing paths and matchings in a similar spirit are addressed in [17].

3 Paths, Cycles, and Beyond

Denote by C_k a cycle of k vertices, D_k a cycle of k vertices triangulated from a vertex, P_k a path of k vertices (that is, of length $k-1$), and S_k a star of k vertices. In addition, $M_{2k} = kP_2$ will stand for any perfect matching on $2k$ vertices. Regarding paths, the following results are known [23].

Theorem E. *If* $k \geq 3$*, then* $2k-3 = R_c(P_k) \leq R_g(P_k) = O(k^{3/2})$.

Thus, the geometric Ramsey number differs from the abstract Ramsey number $R(P_k) = \lfloor 3k/2 \rfloor - 1$ proved by Gerencsér and Gyárfás [14]. The lower bound is implied by the following simple construction. Let G be a complete convex geometric graph on $2k-4$ vertices, and let $p, v_1, v_2, \ldots, v_{k-3}, q, u_1, u_2, \ldots, u_{k-3}$ be its vertices listed in clockwise order. For all i, j, color all edges v_iv_j, pv_i, and qv_i blue; u_iu_j, pu_i, and qu_i red; v_iu_j red if $i+j$ is odd and blue if $i+j$ is even. The edge pq can have any color. It is not difficult to check that this graph contains no noncrossing monochromatic path of length $k-1$. The illustration below shows the blue edges in the case $k=6$.



The upper bound concerning the convex case is a consequence of the following particular case of Theorem D.

Theorem F. *If a convex geometric graph of $n \geq k+1$ vertices has more than $\lfloor (k-1)n/2 \rfloor$ edges, then it contains a noncrossing path of length k.*

If one is satisfied with the weaker bound $(k-1)n$, this result can be obtained using a simple induction on k: Omitting at each vertex the left- and rightmost edge, thus altogether at most $2n$ edges, if the resulting graph contains a noncrossing path of length $k-2$, the original must contain a noncrossing path of length k. The tight bound can be obtained with a more subtle induction argument; see [23].

The estimate $R_c(P_k) \leq 2k-3$ follows immediately from Theorem F when k is odd: Take a 2-colored complete convex geometric graph on $2k-3$ vertices. As the number of edges, $(2k-3)(k-2)$, is odd, the larger color class has more than $\lfloor (2k-3)(k-2)/2 \rfloor$ edges and thus contains a noncrossing path of length $k-1$. The even case can be resolved by a subtle analysis of the proof of Theorem F given in [23]. Most likely, the general upper bound is very far from the truth, but it may be very difficult to find the right order of magnitude.

Problem 4. Improve upon the upper bound $R_g(P_k) = O(k^{3/2})$.

The weaker, but somewhat more general, bound $R_g(P_{k+1}, P_{l+1}) \leq kl+1$ can be proved by the following argument [22]. Let p_i $(0 \leq i \leq kl)$ denote the vertices of a complete geometric graph. Suppose that they are listed in increasing order of their x-coordinates, which are all distinct. Define a partial ordering on the vertex set P as follows. Let $p_i < p_j$ if $i < j$ and there is an x-monotone red path connecting p_i to p_j. By Dilworth's theorem [11], one can find either $k+1$ vertices that form a totally ordered subset $Q \subset P$, or $l+1$ vertices that are pairwise incomparable. In the first case, there is an x-monotone red path visiting every vertex of Q. In the second case, there is an x-monotone blue path of length l, because any two incomparable elements are connected by a blue edge. The bound follows, noting that an x-monotone path cannot intersect itself.

The situation changes dramatically if instead of paths we consider cycles. Following up on the work of Chartrand and Schuster [7] and Bondy and Erdős [3], the Ramsey numbers $R(C_k, C_\ell)$ were determined completely by Faudree and Schelp [12] and independently by Rosta [30]; see also [24]. For example, $R(C_k) = 3k/2-1$ for k even and $R(C_k) = 2k-1$ for odd values of $k > 3$. In contrast, $R_c(C_k) > (k-1)^2$ was observed in [23], and independently by Harborth and Lefmann [20]. In fact, the following more general estimate is true [24].

Theorem G. $R_c(C_k, P_\ell) \geq (k-1)(\ell-1)+1$.

Indeed, construct a complete convex geometric graph on $(k-1)(\ell-1)$ vertices by choosing $k-1$ pairwise disjoint arcs on a circle and placing $\ell-1$ different points on each arc. Color all the edges connecting two vertices on the same arc blue and all the remaining edges red. Obviously, there is no noncrossing blue path with ℓ vertices, and it is easy to check that the graph does not contain a noncrossing red cycle of length k either. On the other hand, the idea explained in proving a quadratic upper bound for the geometric Ramsey number of paths can be used to show that these results are best possible up to a constant multiplicative factor [23, 25].

Theorem H. *$R_g(D_k, D_\ell) \le (k-2)(\ell-1)+(k-1)(\ell-2)+2$ holds for arbitrary integers $k, \ell \ge 3$.*

Since $R_c(C_4) = 14$ [4], this bound is tight for $k = \ell = 4$. Given that $R_c(C_3, C_\ell) = 3\ell - 3$ holds for every $\ell \ge 3$ [25], it is also tight for the case $k = 3$. Besides these cases, nothing beyond Theorem G is known. Let $X, Y \in \{C, D, P\}$ (except $X = Y = P$).

Problem 5. Find the exact values of any of the functions $R_c(X_k, Y_\ell)$, $R_g(X_k, Y_\ell)$.

The estimate $R_c(C_k, G_\ell) \ge (k-1)(\ell-1)+1$ extends to any connected outerplanar graph G_ℓ on ℓ vertices [24]. The following old result of Chvatal [8] can be used to obtain a matching upper bound in certain cases.

Theorem I. *$R(K_k, T_\ell) = (k-1)(\ell-1)+1$ holds for any tree T_ℓ on ℓ vertices.*

This, together with Theorem A, implies $R_g(H_k, S_\ell) \le (k-1)(\ell-1)+1$ for every outerplanar graph H_k on k vertices. Since $R_g(H_k, P_\ell) \le (k-1)(\ell-1)+1$ can also be proved using the above-explained argument, it is quite plausible that similar estimates hold for any tree T_ℓ on ℓ vertices.

Problem 6. Is it true that $R_c(C_k, T_\ell) = R_g(C_k, T_\ell) = (k-1)(\ell-1)+1$ holds for any tree T_ℓ on ℓ vertices?

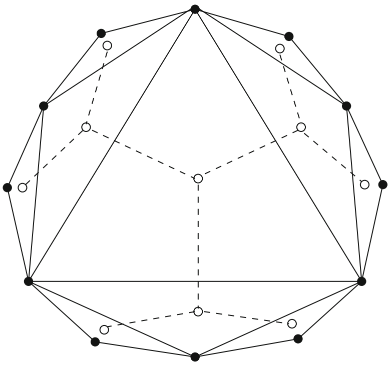
Probably this is true even if C_k is replaced by any outerplanar graph H_k on k vertices, which contains a Hamiltonian cycle. Even though the Ramsey function $R(n)$ is exponentially large, it may well be that all geometric Ramsey numbers are relatively small.

Problem 7. Is there a polynomial f such that $R_g(G_n) < f(n)$ holds for every outerplanar graph G_n with n vertices?

It is even possible that $R_g(G_n) < cn^2$ holds with a universal constant c. Some very recent results pointing this direction are the following. The 'ladder' graph L_{2n} is obtained by connecting vertex disjoint paths $p_1 \dots p_n$ and $q_1 \dots q_n$ with the edges $p_i q_i$ for $1 \le i \le n$. Then $R_c(L_{2n}) = O(n^3)$ and $R_g(L_{2n}) = O(n^{10})$ [9]. Moreover, Cibulka et al. [9] also obtained the polynomial bound $R_c(G_n) = O(n^9)$ for every triangulated n-gon G_n whose weak dual (obtained from the dual graph by removing the vertex corresponding to the infinite face) is a path. It would be very interesting to see if the above conjecture is true for the family of graphs $G_{3 \cdot 2^n}$ obtained from a single triangle by gluing triangles to the outer edges in n iterations. The figure shows the graph G_{12} together with its weak dual, which is a binary tree rooted at the initial triangle.

4 Matchings

The off-diagonal Ramsey number for paths determined by Gerencsér and Gyárfás [14] implies that $R(M_{2k}, M_{2\ell}) \le k + 2\ell - 1$ holds for $k \le \ell$; this also follows from a more general statement due to Cockayne and Lorimer explained below. The result



is sharp, as one can readily check by taking $K_{k+2\ell-2}$ and coloring the edges of a complete subgraph on $2\ell-1$ vertices blue, and all the remaining edges red. The following theorem first proved in [22] states that this Ramsey number does not change by adding the geometric constraint.

Theorem J. $R_{\text{g}}(M_{2k}, M_{2\ell}) = k + \ell + \max\{k, \ell\} - 1$.

In particular, $R_{\text{c}}(M_{2k}) = R_{\text{g}}(M_{2k}) = 3k-1$. For the convex case, it follows by a trivial induction. For the general case, there is also a proof [21] that can be turned into an $O(k^{\log\log k+2})$-time algorithm, which finds k pairwise disjoint edges of the same color in a bi-colored complete convex geometric graph on $3k-1$ vertices. A similar statement can be made regarding the proof of Theorem C.

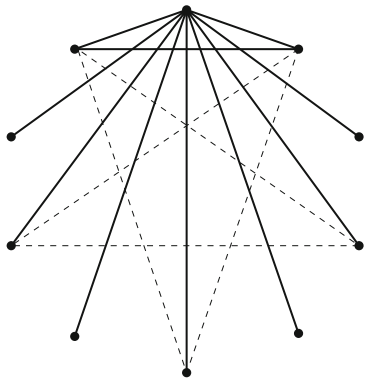
The multicolor Ramsey number of matchings was determined by Cockayne and Lorimer [10] as

$$R(M_{2k_1}, M_{2k_2}, \ldots, M_{2k_t}) = \sum_{i=1}^{t} k_i + \max_{1\le i\le t} k_i - t + 1.$$

When it comes to matchings, the idea of merging color classes yields reasonable upper bounds for multicolor Ramsey numbers. Thus, for the diagonal Ramsey number $R_{\text{g}}^{(t)}(M_{2k}) = R_{\text{g}}(\underbrace{M_{2k}, \ldots, M_{2k}}_{t \text{ times}})$, Theorem J implies the general upper bound

$$R_{\text{c}}^{(t)}(M_{2k}) \le R_{\text{g}}^{(t)}(M_{2k}) \le \begin{cases} \frac{3t}{2}k - \frac{3t}{2} + 2 & \text{for } t \text{ even,} \\ \frac{3t+1}{2}k - \frac{3t+1}{2} + 2 & \text{for } t \text{ odd.} \end{cases}$$

This upper bound is sharp also for $t = 4$, but most likely it is not for larger values of t. If $t \ge 2$ and $k \ge 6t - 10$, then $R_{\text{c}}^{(t)}(M_{2k}) \ge (6/5)tk$. See [25] for the proof of these results. It is also true that the upper bound is sharp for $t = 6$ and $k = 2$. Consider 10 points arranged at the vertices of a regular 10-gon and color the line segments connecting them using six colors as follows. For the first color class, take the edges shown in the figure below. Four additional classes are obtained by rotating the first at the angle $2\pi i/5$ around the center for $1 \le i \le 4$. The edges not yet colored form a pentagram (dashed lines), which does not have two pairwise disjoint edges either.



But for larger even values of t, the bound, at least in the convex case, is not sharp anymore. However, despite repeated efforts, nothing other than the following is known, except that one can also obtain a similar improvement for t odd.

Theorem K. *Let $t = 2s$ be an even integer, $s \ge 4$. Then $R_c^{(t)}(M_4) \le 3s+1$.*

Proof. Let $s > 3$. Take a complete convex geometric graph G_0 on $3s+1$ vertices and color the edges using $2s$ different colors. Assume, for contradiction, that any two edges of the same color have a point in common. There must be two boundary edges of the same color $\mathcal{C}$. By our assumption, these two edges share a vertex; thus, they must be of the form pq and qr for three consecutive vertices p,q,r. Now, except the edge pr, any edge of color $\mathcal{C}$ must have q as an endpoint. Consider the graph G_1 obtained from G_0 by removing the edge pr as well as the vertex q and all edges incident to it. This graph is a complete convex geometric graph on $3s$ vertices, with one boundary edge, namely, pr missing. Its edges are colored with $2s-1$ different colors. Since it has $3s-1$ remaining boundary edges, two of them must have the same color, and we may repeat the above argument.

After the ith iteration, we obtain a graph G_i, which is a complete convex geometric graph on $3s+1-i$ vertices, with i boundary edges missing, whose edges are colored with $2s-i$ colors. The next iterative step can be executed if the number of remaining boundary edges, that is, $3s+1-2i$, is larger than the number of colors. Thus, eventually we obtain G_{s+1}, a complete convex geometric graph on $2s$ vertices, with $s+1$ boundary edges missing, whose edges are colored with $s-1$ colors. The number of remaining boundary edges is also $s-1$. Now we distinguish between two cases.

If two boundary edges of the same color still remain, we can construct the graph G_{s+2}. It has $(2s-1)(s-1)-(s+2)$ edges, colored with $s-2$ colors. In each color class, any two edges have a point in common. It is folklore that the size of such a class cannot exceed the number of vertices, which is $2s-1$. Thus, the total number of edges cannot exceed $(s-2)(2s-1)$. Given that $s>3$, we have $(2s-1)(s-1)-(s+2)>(s-2)(2s-1)$, a contradiction.

Otherwise, we have exactly one boundary edge of each color class, missing at least two vertices. If there is an edge of G_{s+1} connecting two missing vertices, then we are done, because such an edge is disjoint from all boundary edges and must fall in the same color class with one of them. This leaves us with the following possibility: There are only two missing vertices, which must be consecutive vertices. Let the vertices of G_{s+1} be listed in clockwise order around the boundary as $v_1, v_2, \ldots, v_{2s}$, with v_3v_4, $v_5v_6, \ldots, v_{2s-1}v_{2s}$ as the remaining boundary edges. Since $s>3$, there is the boundary edge v_7v_8 among them; thus, both v_2v_5 and v_1v_6 are edges of G_{s+1}. Because these edges are disjoint from each boundary edge except for v_5v_6, they must have the same color as v_5v_6, but this is impossible, since v_2v_5 and v_1v_6 do not have a common point. □

It may be difficult to determine the exact values of these Ramsey numbers, but probably with some additional idea this argument can be extended to improve the upper bound by a multiplicative factor instead of just by the additive constant 1. It is also plausible that the general upper bound can be improved upon considerably.

Problem 8. Is there a constant $\alpha<3/2$ such that $R_{\mathrm{c}}^{(t)}(M_{2k})<\alpha tk$ holds for $t\geq t_0$, $k\geq k_0$?

An additional motivation for the investigation of this question has been a problem studied by Araujo et al.[1]. The *convex segment disjointness graph* D_n is obtained from the complete convex geometric graph G_n; its vertices are the edges of G_n, two such vertices being connected if the line segments they represent are disjoint. Based on Theorem J, the lower bound $\chi(D_n)\geq 2\lfloor\frac{n+1}{3}\rfloor-1$ was obtained in [1]. Any progress on Problem 7 would have yielded an improved lower bound. However, Fabila-Monroy and Wood [13] recently proved the very strong lower bound $\chi(D_n)\geq n-\sqrt{2n+\frac{1}{4}}+\frac{1}{2}$, somewhat reducing the interest in Problem 8.

5 Ramsey Multiplicity

Let G denote a simple graph without isolated vertices. The Ramsey multiplicity $RM(G)$ of G is the minimum number of monochromatic copies of G in any 2-coloring of the edges of $K_{R(G)}$. The concept was introduced by Harary and Prins in [19], who made the conjecture that $R(G)$ is usually large and attains the smallest possible value 1 if and only if G is a star on m vertices, where $m=2$ or $m>1$ is an odd integer. That is, when a monochromatic copy of G must occur in any 2-coloring of the edges of K_n, it should occur many times. For example, for the class of odd cycles C_n, an exponential lower bound on $f(n)=RM(C_n)$ was obtained by Rosta and Surányi [31], and in fact it is known to be superexponential [26]:

$$\mathrm{e}^{(\frac{1}{24}-o(1))n\log n}<RM(C_n)<\mathrm{e}^{n\log n}.$$

This concept can be introduced to the geometric graph Ramsey theory we investigate in this chapter, for which we will use the notations $RM_c(G)$ and $RM_g(G)$. Apart from trivial cases (like stars, for example), these multiplicities are expected to be much smaller than $RM(G)$. According to the Gerencsér–Gyárfás theorem and Theorem J, we have $R_g(M_{2n}) = R_c(M_{2n}) = R(M_{2n}) = 3n-1$. It is an easy exercise to prove a superexponential lower bound of the form $e^{cn\log n}$ for $RM(M_{2n})$, but the situation is different in the geometric setting.

Observation L. *$RM_g(M_{2n}) \leq RM_c(M_{2n}) \leq \binom{2n-1}{n} < 4^n$ holds for every positive integer n.*

Proof. Take a complete convex geometric graph on $3n-1$ vertices, which are listed in clockwise order as $v_1, v_2, \dots, v_{3n-1}$. Color the edges of a complete subgraph induced by the vertex set $\{v_{n+1}, \dots, v_{3n-1}\}$ blue, and all the remaining edges red. Obviously, the graph does not contain M_{2n} as a blue monochromatic subgraph. On the other hand, every edge in a monochromatic red M_{2n} must have an endpoint in $\{v_1, v_2, \dots, v_n\}$, and all these endpoints must be different. Consequently, the other endpoints must fall in the set $\{v_{n+1}, v_{n+2}, \dots, v_{3n-1}\}$. Moreover, if $\{v_1v_{i_1}, v_2v_{i_2}, \dots, v_nv_{i_n}\}$ is a noncrossing matching, then $(i_1, i_2, \dots, i_n)$ must be a strictly decreasing sequence, hence the bound. □

We investigated several other natural candidates for a better bound, but we were not able to come up with a construction that would yield a subexponential bound. So we think that the following may be true.

Problem 9. Is there a constant $\alpha > 1$ such that $RM_c(M_{2n}) = 1$?

Probably the same is true for $RM_g(M_{2n})$. The family of paths is another nontrivial example where the exact values of the (convex) geometric Ramsey numbers are known, although they are essentially larger than the corresponding abstract Ramsey numbers. Therefore, we feel that the following problem is not hopeless.

Problem 10. Find reasonable estimates for $RM_c(P_n)$.

Acknowledgements Part of this paper was written during the special semester on Discrete and Computational Geometry held at the EPFL Lausanne, sponsored by the Centre Interfacultaire Bernoulli and the Swiss National Science Foundation. This work was also partially supported by the ESF EUROCORES programme EuroGIGA, CRP Graph Drawing, and Hungarian Scientific Research Grant OTKA NN102029.

References

1. G. Araujo, A. Dumitrescu, F. Hurtado, M. Noy, J. Urrutia, On the chromatic number of some geometric Kneser graphs. Comput. Geom. Theor. Appl. **32**, 59–69 (2005)
2. A. Bialostocki, P. Dierker, W. Voxman, Either a graph or its complement is connected: A continuing saga. Manuscript (2001)

3. J.A. Bondy, P. Erdős, Ramsey numbers for cycles in graphs. J. Combin. Theor. B **14**, 46–54 (1973)
4. A. Bialostocki, H. Harborth, Ramsey colorings for diagonals of convex polygons. Abh. Braunschweig. Wiss. Ges. **47**, 159–163 (1996)
5. P. Brass, Gy. Károlyi, P. Valtr, A Turán-type extremal theory of convex geometric graphs, in *Discrete and Computational Geometry*, ed. by B. Aronov et al., Algorithms and Combinatorics, vol. 25 (Springer, Berlin, 2003), pp. 275–300
6. S.A. Burr, Either a graph or its complement contains a spanning broom. Manuscript (1992)
7. G. Chartrand, S. Schuster, On a variation of the Ramsey number. Trans. Am. Math. Soc. **173**, 353–362 (1972)
8. V. Chvátal, Tree—complete graph Ramsey numbers. J. Graph Theor. **1**, 93 (1977)
9. J. Cibulka, P. Gao, M. Krcal, T. Valla, P. Valtr, Geometric Ramsey theorems for pathwidth-2 outerplanar triangulations. Manuscript (2011)
10. E.J. Cockayne, P.J. Lorimer, The Ramsey number for stripes. J. Austral. Math. Soc. A **19**, 252–256 (1975)
11. R.P. Dilworth, A decomposition theorem for partially ordered sets. Ann. Math. **51**, 161–166 (1950)
12. R.J. Faudree, R.H. Schelp, All Ramsey numbers for cycles in graphs. Discrete Math. **8**, 313–329 (1974)
13. R. Fabila-Monroy, D.R. Wood, The chromatic number of the convex segment disjointness graph, in *XIV Spanish Meeting on Computational Geometry, CRM Documents*, vol. 8, (CRM, Bellaterra, 2011), pp. 47–50
14. L. Gerencsér, A. Gyárfás, On Ramsey-type problems. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. **10**, 167–170 (1967)
15. P. Gritzmann, B. Mohar, J. Pach, R. Pollack, Embedding a planar triangulation with vertices at specified points (solution to problem E3341). Am. Math. Month. **98**, 165–166 (1991)
16. A. Gyárfás, Partition coverings and blocking sets in hypergraphs [in Hungarian]. Commun. Comput. Autom. Inst. Hungar. Acad. Sci. No. 71, (1977), 62 pages
17. A. Gyárfás, Ramsey and Turán-type problems for non-crossing subgraphs of bipartite geometric graphs. Ann. Univ. Sci. Budapest. Eötvös Sect. Math. to **54**, 47–56 (2011)
18. A. Gyárfás, Large monochromatic components in edge colorings of graphs: A survey, in *Ramsey Theory, Progress in Mathematics*, vol. 285 (Springer, Berlin, 2011), pp. 77–96
19. F. Harary, G. Prins, Generalized Ramsey theory for graphs. IV. The Ramsey multiplicity of a graph. Networks **4**, 163–173 (1974)
20. H. Harborth, H. Lefmann, Coloring arcs of convex sets. Discrete Math. **220**, 107–117 (2000)
21. Gy. Károlyi, J. Pach, G. Tardos, G. Tóth, An algorithm for finding many disjoint monochromatic edges in a complete 2-colored geometric graph, in *Intuitive Geometry*, ed. by I. Bárány and K. Böröczky, Bolyai Soc. Math. Studies, vol. 6 (J. Bolyai Math. Society, Budapest, 1997), pp. 367–372
22. Gy. Károlyi, J. Pach, G. Tóth, Ramsey–type results for geometric graphs. I. Discrete Comput. Geom. **18**, 247–255 (1997)
23. Gy. Károlyi, J. Pach, G. Tóth, P. Valtr, Ramsey–type results for geometric graphs. II. Discrete Comput. Geom. **20**, 375–388 (1998)
24. Gy. Károlyi, V. Rosta, Generalized and geometric Ramsey numbers. Theoret. Comput. Sci. **263**, 87–98 (2001)
25. Gy. Károlyi, V. Rosta, On geometric graph Ramsey numbers. Graphs Combinator. **25**, 351–363 (2009)
26. Gy. Károlyi, V. Rosta, On the Ramsey multiplicity of the odd cycles. J. Graph Theor. submitted (2011)
27. C. Keller, M.A. Perles, On the smallest sets blocking simple perfect matchings in a convex geometric graph. Israel J. Math., **187**, 465–484 (2012)

28. W. Moser, J. Pach, Recent developments in combinatorial geometry, in *New Trends in Discrete and Computational Geometry*, ed. by J. Pach, Algorithms and Combinatorics, vol. 10 (Springer-Verlag, New York, 1993), pp. 281–302
29. F.P. Ramsey, On a problem of formal logic. Proc. Lond. Math. Soc. **30**, 264–286 (1930)
30. V. Rosta, On a Ramsey type problem of J.A. Bondy and P. Erdős. I–II. J. Combin. Theor. B **15**, 94–120 (1973)
31. V. Rosta, L. Surányi, A note on the Ramsey multiplicity of the circuit. Period. Math. Hungar. **7**, 223–227 (1976)

Blockers for Noncrossing Spanning Trees in Complete Geometric Graphs

Chaya Keller, Micha A. Perles, Eduardo Rivera-Campo, and Virginia Urrutia-Galicia

Abstract In this chapter, we present a complete characterization of the smallest sets that block all the simple spanning trees (SSTs) in a complete geometric graph. We also show that if a subgraph is a blocker for all SSTs of diameter at most 4, then it must block all simple spanning subgraphs and, in particular, all SSTs. For convex geometric graphs, we obtain an even stronger result: Being a blocker for all SSTs of diameter at most 3 is already sufficient for blocking all simple spanning subgraphs.

1 Introduction

A *geometric graph* is a graph whose vertices are points in general position in the plane[1] and whose edges are segments connecting pairs of vertices. Let $G = (V(G), E(G))$ be a complete geometric graph and let $\mathcal{F}$ be a family of subgraphs of G. We say that a subgraph B of G *blocks* $\mathcal{F}$ if it has at least one edge in common with each member of $\mathcal{F}$. We denote by $\mathcal{B}(\mathcal{F})$ the collection of all smallest (i.e., having the smallest possible number of edges) subgraphs of G that block $\mathcal{F}$, and we call its elements *blockers* of $\mathcal{F}$.

[1]Formally, the assumption is that an edge never contains a vertex in its relative interior. In the case of a complete geometric graph, which we consider in this chapter, this implies that the vertices are in general position (i.e., that no three vertices lie on the same line).

C. Keller (✉) • M.A. Perles
Einstein Institute of Mathematics, Hebrew University Jerusalem 91904, Israel
e-mail: ckeller@math.huji.ac.il; perles@math.huji.ac.il

E. Rivera-Campo • V. Urrutia-Galicia
Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa Av. San Rafael Atlixco 186, México D.F. 09340, Mexico
e-mail: erc@xanum.uam.mx; vug@xanum.uam.mx

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_20,

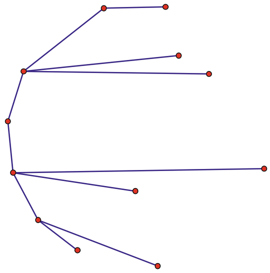


Fig. 1 A comb in a nonconvex geometric graph on 12 vertices

Blockers for several families of subgraphs were studied in previous papers. For example, the set $\mathcal{B}(SPM)$ of blockers for the family of all simple (i.e., noncrossing) perfect matchings in a complete convex geometric graph of even order was characterized in [5], and the family of corresponding co-blockers (i.e., $\mathcal{B}(\mathcal{B}(SPM))$) was characterized in [6]. The characterizations give rise to interesting structures, such as classes of *caterpillars* [1, 2].

In this chapter, we study the set $\mathcal{B}(SST)$ of blockers for the family of simple spanning trees (SSTs) of a complete geometric graph, and we give the following characterization.

Definition 1.1. A simple spanning subgraph B of a complete geometric graph G is a *comb* of G if

1. The intersection of B with the boundary of $\text{conv}(G)$ is a simple path P. We call this path the *spine* of B.
2. Each vertex in $V(G) \setminus P$ is connected by a unique edge to an interior vertex of P.
3. For each edge e of B, the line $l(e)$ spanned by e does not cross any edge of B.[2]

Note that a comb B, regarded as an abstract tree, is a caterpillar and that the derived tree is the path P with the first and last edges removed. An example of a comb is shown in Fig. 1.

Theorem 1.2. *A graph B is a blocker for the family of all simple spanning trees of a complete geometric graph G if and only if B is either a star (i.e., the set of all edges in G that emanate from a single vertex) or a comb of G.*

We note that in the convex case, this characterization can be derived by combining a result of Hernando [3] that characterizes those SSTs that meet all other SSTs, with a result of Károlyi et al. [4] that shows that any 2-coloring of a complete geometric graph contains a monochromatic SST. Theorem 1.2 was recently used in [7] to show that if G is a complete geometric graph with n vertices in which

[2] Note that the line $l(e)$ avoids all vertices of G except the endpoints of e, since $V(G)$ is in general position.

exactly one vertex does not lie on the boundary of conv(G), and c is a coloring of the edges of G with $n(n-1)/2-n+1$ colors, then G has a simple spanning tree all of whose edges have different colors.
We also present several refinements of Theorem 1.2.

For a complete geometric graph G and for $k \in \mathbb{N}$, denote by $\mathcal{T}_{\leq k}(G)$ the family of all simple spanning trees of G with diameter at most k.

Theorem 1.3. *Let B be a subgraph of a complete geometric graph G. If $B \in \mathcal{B}(\mathcal{T}_{\leq 4}(G))$, then B is either a star or a comb of G.*

In the case of complete convex geometric graphs, we can replace diameter 4 by diameter 3, as follows:

Theorem 1.4. *Let B be a subgraph of a complete convex geometric graph G. If $B \in \mathcal{B}(\mathcal{T}_{\leq 3}(G))$, then B is a comb of G.*

The two latter results improve Theorem 1.2 by showing that being a blocker for SSTs of diameter at most 4 (or even at most 3 in the convex case) is sufficient for being a blocker for all SSTs. These results are tight in the sense that $\mathcal{T}_{\leq 4}(G)$ cannot be replaced by $\mathcal{T}_{\leq 3}(G)$ in Theorem 1.3, as we show by an example in Sect. 5, and $\mathcal{T}_{\leq 3}(G)$ cannot be replaced by $\mathcal{T}_{\leq 2}(G)$ in Theorem 1.4, since any spanning subgraph blocks all trees in $\mathcal{T}_{\leq 2}(G)$ but not all SSTs of G.

Finally, the following result improves Theorem 1.2 in the opposite direction. We say that $H \subset G$ is a *simple spanning subgraph* (SSS) of G if H is noncrossing and has no isolated vertices; i.e., every vertex of G is incident to an edge of H.

Theorem 1.5. *Let B be a subgraph of a complete geometric graph G. If B is a star or a comb of G, then B blocks all simple spanning subgraphs of G.*

The chapter is organized as follows: In Sect. 2, we give precise definitions and notations used throughout the chapter. In Sect. 3, we prove properties of blockers for $\mathcal{T}_{\leq 3}(G)$ common to the general case and the convex case. In Sects. 4 and 5, we prove Theorems 1.4 and 1.3, respectively; and in Sect. 6, we complete the proof of Theorem 1.2 and prove Theorem 1.5 by showing that $\mathcal{B} \subset \mathcal{B}(SSS)$ for any complete geometric graph G, where $\mathcal{B}$ denotes the family of all combs of G.

2 Definitions and Notations

In this section, we present some definitions and notations used in the chapter.

Geometric graphs. Throughout the chapter, G is a complete geometric graph on n vertices. The sets of vertices and edges of G are denoted by $V(G)$ and $E(G)$, respectively. The convex hull of $V(G)$ is denoted by conv(G). Vertices in $V(G)$ and edges in $E(G)$ that lie on the boundary of conv(G) are called *boundary vertices* and *boundary edges* of G, respectively. A geometric graph is *simple* if it does not contain a pair of crossing edges. For more information on geometric graphs, the reader is referred to [8].

Caterpillars. Throughout the chapter, T is a tree. A tree T is a *caterpillar* if the derived graph T' (i.e., the graph obtained from T by removing all leaves and their incident leaf edges) is a path (or is empty). A longest path in a caterpillar T is called a *spine* of T. (Note that any edge of T either belongs to every spine or is a leaf edge of T.) If the diameter of T is 3, then T contains an edge $[x,y]$ such that each vertex in $V(T)$ is at distance at most 1 from either x or y. Such an edge $[x,y]$ is called the *central edge* of T. (Note that any tree of diameter 3 is a caterpillar.)

General notations in the plane. Throughout the chapter, l denotes a line. Each line l partitions the plane into open half-planes. We denote them by l^+ and l^- and call them the *sides* of the line. The unique line that contains two points $a,b \in \mathbb{R}^2$ is denoted by $l(a,b)$. The complement of a set A in $\mathbb{R}^2$, i.e., the set $\mathbb{R}^2 \setminus A$, is denoted by A^c.

3 Some Properties of Blockers for SSTs of Diameter at Most 3

In this section, we establish several properties of blockers for SSTs of diameter at most 3. First, we show that the number of edges in these blockers is $n-1$ (where n is the number of vertices in G), and then we show that all such blockers are caterpillars.

3.1 The Size of the Blockers

Proposition 3.1. *Let G be a complete geometric graph on n vertices. Then the size (i.e., number of edges) of the blockers for SSTs of diameter at most* 3 *in G is $n-1$.*

As any star blocks all SSTs of diameter at most 3 (and actually even all spanning subgraphs), the size of the blockers is at most $n-1$. The other inequality is a consequence of the following unpublished result of Perles (unpublished, 1987).

Theorem 3.2. *Let G_1 be a geometric graph on n vertices. If $|E(\overline{G_1})| \leq n-2$, then G_1 includes an SST of diameter at most* 3. *($\overline{G_1}$ denotes the graph complementary to G_1, on the same set of vertices.)*

Theorem 3.2 implies that a set of at most $n-2$ edges cannot block all the SSTs of diameter at most 3, since its complement includes such an SST. Thus, the size of blockers is at least $n-1$, which completes the proof of Proposition 3.1. For the sake of completeness, we present here the proof of Theorem 3.2.

Proof. Since $|E(\overline{G_1})| \leq n-2$, the following two statements hold:

1. $\overline{G_1}$ has a vertex of degree 0 or 1.
2. $\overline{G_1}$ is not connected.

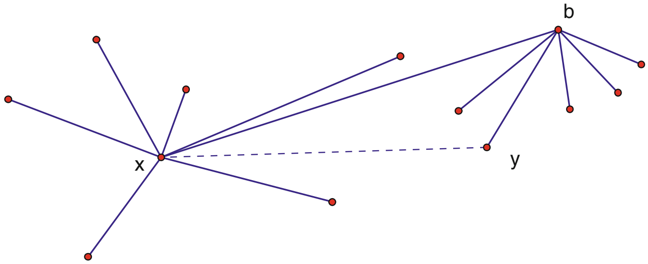


Fig. 2 An SST of diameter 3 that avoids $\overline{G_1}$

Since the number of edges in $\overline{G_1}$ is smaller than the number of vertices, at least one connected component of $\overline{G_1}$ is a tree. Denote one such component by A. If A is a single vertex x', then G_1 contains the star centered at x', which is an SST of diameter 2 (assuming $n \geq 2$). Otherwise, A has a leaf x. Denote the (only) neighbor of x in A by y.

Consider the ray $\vec{xy}$, and turn it around x until it hits a vertex $b \notin A$ for the first time. [There must be a vertex $b \notin A$, since $\overline{G_1}$ is not connected, and thus $A \neq V(G)$. Moreover, b is unique, since a ray emanating from x cannot contain two other vertices of G.] Denote by H the closed convex cone bounded by the rays $\vec{xy}$ and $\vec{xb}$.

Let T be the subgraph of G whose edges are

$$\{[x,z] : z \in V(G) \cap H^c\} \cup \{[b,w] : w \in (V(G) \setminus \{b\}) \cap H\},$$

as illustrated in Fig. 2. It is clear by the construction that T is an SST of diameter 3 (with central edge $[x,b]$), or of diameter 2 (if $V(G) \subset H$). We claim that $T \subset G_1$. Indeed, the edges $[x,z]$, where $z \in H^c \cap V(G)$ are all in G_1, as the only edge in $\overline{G_1}$ that contains x is $[x,y]$. The edges $[b,w]$, where $w \in H \cap (V(G) \setminus \{b\})$, are also in G_1, since the vertices $\{w : w \in (V(G) \setminus \{b\}) \cap H\}$ belong to A, whereas b belongs to another connected component of $\overline{G_1}$. Therefore, $T \subset G_1$, which completes the proof.

We note that the proof of Theorem 3.2 implies a stronger statement.

Proposition 3.3. *Let G be a complete geometric graph on n vertices. Then the blockers for SSTs of diameter at most* 3 *in G are spanning trees.*

Proof. Let B be a blocker for SSTs of diameter at most 3 in G. By Proposition 3.1, $|E(B)| = n - 1$. It is clear that B is a spanning subgraph of G without isolated vertices, since otherwise it avoids a star, which is an SST of diameter 2. If B is not a tree, then the two statements at the beginning of the proof of Theorem 3.2 clearly hold (i.e., B has a vertex of degree 0 or 1 and is not connected). Thus, by the proof of Theorem 3.2, $\bar{B}$ contains an SST of diameter at most 3, a contradiction. □

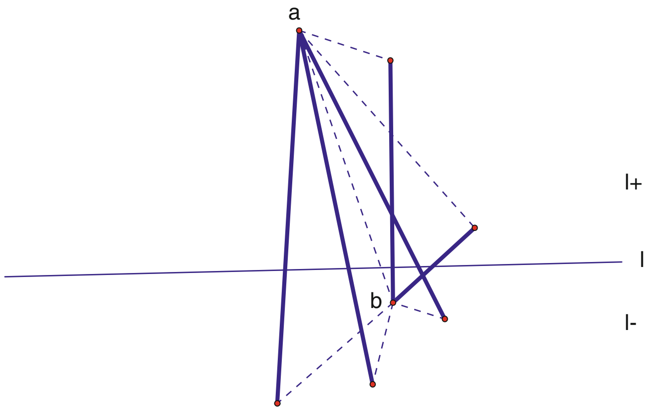


Fig. 3 An illustration of the proof of Observation 3.4. The edges of B are depicted by *heavy lines*, and the edges of T by *dotted lines*

3.2 The Blockers Are Caterpillars

In order to further characterize the blockers, we use two observations.

Observation 3.4. *Let B be a subgraph of G. Assume there exist two vertices $a,b \in V(G)$ and a line l such that*

1. $[a,b] \notin E(B)$, *and*
2. *a and all neighbors of b in B lie on one (open) side l^+ of l, while b and all neighbors of a in B lie on the other (open) side l^- of l.*

Then $B \notin \mathcal{B}(\mathcal{T}_{\leq 3})$.

Proof. If conditions (1) and (2) hold, then B avoids the following SST, as illustrated in Fig. 3: $T = (V(G), E(T))$, where

$$E(T) = \{[a,b]\} \cup \{[a,x] : x \in V(G) \cap (l^+ \cup l), x \neq a\} \cup \{[b,y] : y \in V(G) \cap l^-, y \neq b\}.$$

It is clear that $\mathrm{diam}(T) \leq 3$ [as the distance (in T) of all vertices from the edge $[a,b]$ is at most 1], that T is crossing-free, and that T avoids B. The assertion follows.

Remark 3.5. It is clear that the observation holds also if a and b have neighbors on l, as long as they do not have a common neighbor on l.

Corollary 3.6. *Assume $B \in \mathcal{B}(\mathcal{T}_{\leq 3})$, and let a,b be two leaves of B. Let the corresponding leaf edges be $[a,c]$ and $[b,d]$. If a,b,c,d are mutually distinct, then the points a,b,d,c (in this order) are the vertices of a convex quadrilateral.*

Proof. Recall that a,b,c,d are vertices of G and therefore are in general position. If they are not in convex position, then the segments $[a,d]$ and $[b,c]$ are disjoint.

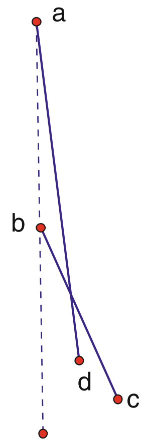


Fig. 4 An illustration to the proof of Observation 3.7

The same holds if they form a convex quadrilateral in a different order, say a,c,b,d or a,b,c,d. If $[a,d]\cap[b,c]=\emptyset$, then these two segments can be separated by a line l. This means that the conditions of Observation 3.4 hold for a,b, and l, and thus $B\notin\mathcal{B}(\mathcal{T}_{\leq 3})$, a contradiction. □

Observation 3.7. *Suppose $B\in\mathcal{B}(\mathcal{T}_{\leq 3})$. Let a be a leaf of B with leaf edge $[a,b]$. Let $c\in V(G)\cap l(a,b)^+$ be the vertex for which the angle $\angle abc$ is maximal [among all vertices in $l(a,b)^+$]. Then $[b,c]\in B$.*

Proof. If $[b,c]\notin B$, then B avoids the following SST:

$$T=\{[a,x]:x\in V(G)\setminus\{a,b\}\}\cup\{[b,c]\}.$$

It is clear that T is a spanning tree of diameter at most 3. T avoids B since the only edge in B that emanates from a is $[a,b]$, and since $[b,c]\notin B$. Finally, T is simple, since if two edges in T cross, then these must be edges of the form $[a,d]$ and $[b,c]$ for some $d\in V(G)$, and in such case, $\angle abd>\angle abc$ (see Fig. 4), contradicting the choice of c. Thus, $B\notin\mathcal{B}(\mathcal{T}_{\leq 3})$, a contradiction. □

Clearly, the same holds for the vertex $c\in V(G)\cap l(a,b)^-$ for which the angle $\angle abc$ is maximal among all vertices in $l(a,b)^-$.

If the leaf edge $[a,b]$ lies on the boundary of conv(G), then the line $l(a,b)$ supports $V(G)$, and thus only one of the sides of $l(a,b)$ [w.l.o.g. $l(a,b)^+$)] contains vertices of G. The vertex $c\in V(G)\cap l(a,b)^+$ for which the angle $\angle abc$ is maximal is the vertex that follows b on the boundary of conv(G), and thus $[b,c]$ is a boundary edge.

If $[a,b]$ is not a boundary edge, then there exist two vertices $c\in V(G)\cap l(a,b)^+$ and $c'\in V(G)\cap l(a,b)^-$ such that the angles $\angle abc,\angle abc'$ are maximal [each with respect to its side of $l(a,b)$], and $[b,c],[b,c']\in B$. This observation is used in the proof of the theorem below.

Now we are ready to present the main result of this section.

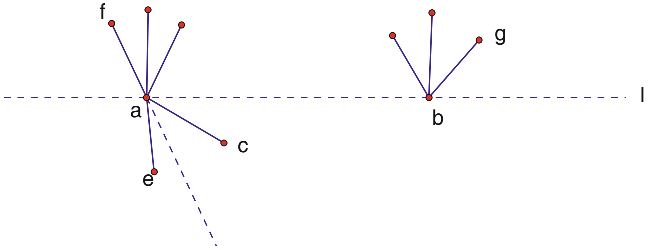


Fig. 5 An illustration to the proof of Theorem 3.8

Theorem 3.8. *Let G be a complete geometric graph. Then any blocker for SSTs of diameter at most* 3 *in G is either a star or a caterpillar with a spine whose terminal edges lie on the boundary of* $\mathrm{conv}(G)$.[3]

Proof. Suppose $B \in \mathcal{B}(\mathcal{T}_{\le 3})$. By Proposition 3.3, B is a tree. If B is a star, we are done. Otherwise, the derived graph B' is a tree with more than one vertex, and thus it has at least two leaves. Let a, b be distinct leaves of B'. By Corollary 3.6 (with a, b playing the role of c, d), all the leaf edges of B that emanate from a and b lie on the same side of the line $l(a,b)$, w.l.o.g., $l(a,b)^+$ (see Fig. 5). We claim that the extremal leaf edges emanating from a and b, denoted in the figure by $[a,f]$ and $[b,g]$, are boundary edges.

Assume on the contrary that $[a,f]$ is not a boundary edge. As described above, it follows from Observation 3.7 that if $c \in l(a,f)^+$ with $\angle fac$ maximal, and $e \in l(a,f)^-$ with $\angle fae$ maximal, then $[a,c],[a,e] \in B$. [Here $l(a,f)^+$ is the side of $l(a,f)$ that contains b.] We have $c \notin l(a,b)^+$, as otherwise $\angle fac < \angle fab$, contradicting the choice of c. Thus, $[a,c]$ is not a leaf edge [since all the leaf edges that emanate from a lie in $l(a,b)^+$]. On the other hand, since a is a leaf of B', all the edges in B that emanate from a except one are leaf edges, and thus, $[a,e]$ is a leaf edge. Therefore, $e \in l(a,b)^+$ (and not as shown in the figure), which contradicts the assumption that $[a,f]$ is the extremal (i.e., the leftmost) leaf edge emanating from a.

So far we have shown that any leaf of B' is contained in a leaf edge of B that is a boundary edge. In order to complete the proof, it suffices to show that B' has only two leaves, which will imply that B' is a path, and hence B is a caterpillar. As the terminal edges of the spine of a caterpillar B emanate from the two leaves of B', it will follow that these edges can be chosen to be boundary edges of G, completing the proof of the theorem.

Assume on the contrary that B' has at least three leaves, say a, b, and c. By the previous steps of the proof, B has leaf edges $[a,d]$, $[b,e]$, and $[c,f]$, which all lie on

[3]Note that usually the term "terminal edges of the spine" of a caterpillar is not defined uniquely. Here and in the sequel, we mean that there exists a spine whose terminal edges are boundary edges, and in all proofs where we consider *the* spine of B, we refer to a particular spine whose terminal edges are boundary edges.

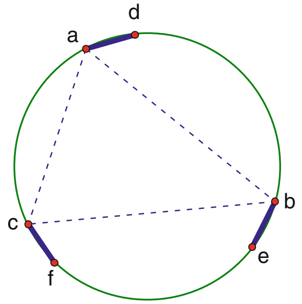


Fig. 6 Three leaf edges on the boundary of conv(G)

the boundary of conv(G). By Corollary 3.6, these edges must satisfy the following three conditions:

- d and e lie on the same side of $l(a,b)$.
- d and f lie on the same side of $l(a,c)$.
- e and f lie on the same side of $l(b,c)$.

But if we somehow orient the boundary of conv(G) (say, counterclockwise), then at least two of the directed edges $\overrightarrow{ad}$, $\overrightarrow{be}$, and $\overrightarrow{cf}$ point the same way (both forward or both backward, as in Fig. 6), and thus fail the condition above. This completes the proof of the theorem. □

4 The Convex Case: Proof of Theorem 1.4

In this section, we assume in addition that G is convex, i.e., that the vertices of G are in convex position in the plane. By Theorem 3.8, a blocker for SSTs of diameter at most 3 in G is a caterpillar, and the terminal edges of its spine lie on the boundary of conv(G). We wish to show that the whole spine of B lies on the boundary of conv(G). This is clear when B is a star. Assume, therefore, that B is not a star. Denote the terminal edges by $[a,a']$ and $[b',b]$, where a and b are the leaves, and $a' \neq b'$, as shown in Fig. 7. [Note that by Corollary 3.6, a and b lie on the same side of the line $l(a',b')$.] Let α and β denote the two closed arcs with endpoints a,b on the boundary of conv(G), as shown in the figure.

Claim 4.1. *If c is a leaf of B, then $c \in \beta$.*

Proof. Assume, w.l.o.g., that $c \neq a$ and $c \neq b$. Denote the leaf edge of B that emanates from c by $[c,c']$. By Corollary 3.6, the following two conditions hold:

- Either $a' = c'$ or a and c lie on the same side of $l(a',c')$.
- Either $b' = c'$ or b and c lie on the same side of $l(b',c')$.

If $c \in \alpha$ (as in the figure), then the two conditions clearly contradict each other. Hence, $c \in \beta$, as claimed. □

Now we are ready to prove the main result of this section.

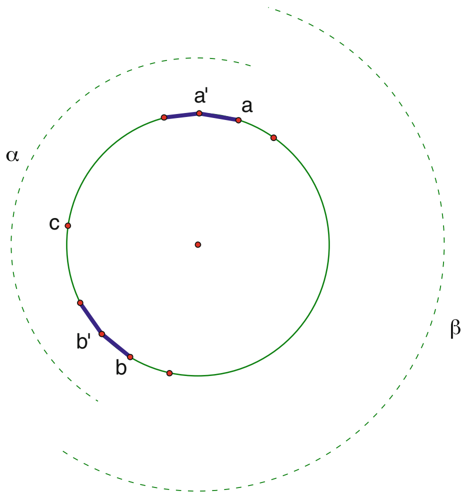


Fig. 7 The terminal edges of the spine of a blocker for SSTs of diameter at most 3 in a complete convex geometric graph divide the boundary into two arcs

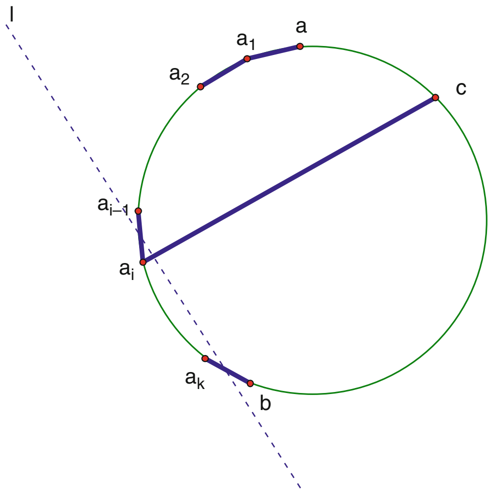


Fig. 8 An illustration of the proof of Theorem 4.2

Theorem 4.2. *Any blocker for SSTs of diameter at most* 3 *in a complete convex geometric graph G is a comb of G.*

Proof. Following the notations of Fig. 7, denote the vertices of G on the arc α by $a = a_0, a' = a_1, a_2, a_3, \ldots, a_k = b', a_{k+1} = b$, as shown in Fig. 8. Since any vertex of a caterpillar that does not belong to its spine is a leaf, it follows from Claim 4.1 that all the vertices $a_0, a_1, \ldots, a_{k+1}$ belong to the spine of B. Let the spine of B be $\langle d_0, d_1, \ldots, d_l, d_{l+1} \rangle$, where $d_0 = a_0 = a$, $d_1 = a_1 = a'$, $d_l = a_k = b'$, and $d_{l+1} = a_{k+1} = b$. If the spine of B is not $\langle a_0, a_1, \ldots, a_k, a_{k+1} \rangle$, then there is a first index ν, $1 \leq \nu \leq l-1$, such that either $d_{\nu+1} \notin \alpha$, or $d_{\nu+1} \in \alpha$, but $d_{\nu+1}$ precedes

d_v on α. Assume $d_v = a_i$, $0 < i < k$. Let l be a line that crosses the segments $[a_{i-1}, a_i]$ and $[a_k, b]$ (see Fig. 8). On one side of l (call it l^-), we have a_i and the only neighbor of b in B, i.e., a_k. On the other side l^+ of l, we have b, all the vertices $a_0, a_1, \dots, a_{i-1}$, and the whole path β. The only neighbors of $a_i(= d_v)$ in B are its predecessor d_{v-1} (which lies in $\{a_0, a_1, \dots, a_{i-1}\}$), its successor d_{v+1} (which lies either in $\{a_0, a_1, \dots, a_{i-1}\}$ or in β), and possibly some leaves (which all lie in β). Thus, all neighbors of a_i in B lie in l^+. Hence, the conditions of Observation 3.4 hold for $d_v = a_i, b$ and the line l, and thus, by the observation, B avoids an SST of diameter at most 3, a contradiction.

Therefore, the spine of B is the boundary path $\langle a_0, a_1, \dots, a_k, b\rangle$. Finally, B is simple, since the only edges in B that can cross are the leaf edges, and these edges do not cross, again by Corollary 3.6. This completes the proof. □

5 The General Case: Proof of Theorem 1.3

In this section, we consider again general geometric graphs. It turns out that in this case, blocking SSTs of diameter at most 3 is not sufficient even for blocking SSTs of diameter 4, as we demonstrate by an example at the end of this section. Thus, we strengthen the assumption and assume now that B is a blocker for all SSTs of diameter at most 4. This allows us to use the following observation.

Lemma 5.1. *Let B be a blocker for SSTs of diameter at most 4 in G. Let b be a boundary vertex of G, and let $[a,b]$ and $[b,c]$ be the two boundary edges of G that contain b. If at least one of these edges is not in B, then b is a leaf of B.*

Proof. Clearly, the degree of b in B is at least 1, as otherwise, B avoids the star centered at b. Assume, on the contrary, that the degree of b in B is at least 2 and that (w.l.o.g.) $[a,b] \notin B$. Let G_1 be the graph obtained from G by omitting the vertex b and all edges that contain it, and let $B_1 = G_1 \cap B$. By the assumption, B_1 is a graph on $n-1$ vertices [where $n = |V(G)|$] that has at most $n-3$ edges. Therefore, by Theorem 3.2, B_1 avoids an SST T_1 of diameter at most 3 in G_1. Since $[a,b]$ is a boundary edge of G, it does not cross any edge of T_1, and thus $T = T_1 \cup \{[a,b]\}$ is an SST of G of diameter at most 4 that avoids B, a contradiction. □

Now we are ready to prove the main result of this section.

Theorem 5.2. *Let G be a complete geometric graph, and let B be a blocker for SSTs of diameter at most 4 in G. Then B is either a star or a comb of G.*

Proof. Assume B is not a star. By Theorem 3.8, B is a caterpillar, and it has a spine $\langle b_0, b_1, \dots, b_{k+1}\rangle$ $(k \geq 2)$ whose extreme edges $[b_0, b_1]$ and $[b_k, b_{k+1}]$ lie on the boundary of conv(G). We would like to show that all the edges $[b_i, b_{i+1}]$ are boundary edges of G. Assume on the contrary that this is not true, and let i, $0 < i < k$, be the smallest index such that $[b_i, b_{i+1}]$ is not a boundary edge. By the assumption,

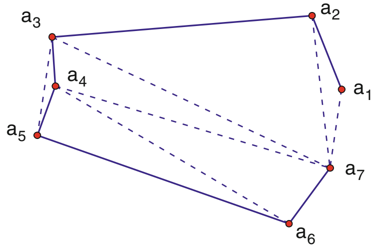


Fig. 9 An example of a subgraph B of a complete geometric graph G that blocks all SSTs of diameter at most 3 but avoids an SST T of diameter 4. The *full lines* represent edges of B, and the *dotted lines* represent edges of T

$[b_{i-1}, b_i]$ is a boundary edge of G. Denote the other boundary edge that contains b_i by $[b_i, c]$. We claim that $[b_i, c] \notin B$. Indeed, as the spine edge $[b_i, b_{i+1}]$ is not a boundary edge, if we had $[b_i, c] \in B$, then this edge would be a leaf edge of B. But this is impossible, since by the proof of Theorem 3.8, B cannot contain three *boundary* leaf edges. Therefore, b_i satisfies the condition of Lemma 5.1, and thus, by that lemma, b_i is a leaf of B, a contradiction.

So far, we have proved that the spine of B lies on the boundary of $\text{conv}(G)$. Consequently, if two edges of B cross, then both must be leaf edges of B. However, a leaf edge of G cannot cross another leaf edge, by Corollary 3.6. Thus, B is simple. Finally, if an edge of G crosses the line spanned by another edge, then there are two possibilities:

1. Both edges are leaf edges. In this case, the convex hull of the union of these two leaf edges is a triangle and not a quadrilateral, contrary to Corollary 3.6.
2. One of the edges is a leaf edge $[b_i, d]$, and the line $l(b_i, d)$ crosses the boundary edge $[b_j, b_{j+1}]$, for some i, j. Assume, w.l.o.g., that $k > j > i$ (and thus, in particular, $i < k - 1$). Consider the edges $[b_i, d]$ and $[b_k, b_{k+1}]$. Both are leaf edges of B, and they lie on different sides of the line $l(b_i, b_k)$. This contradicts Corollary 3.6, since $B \in \mathcal{B}(\mathcal{T}_{\leq 4}) \subset \mathcal{B}(\mathcal{T}_{\leq 3})$.

This completes the proof of the theorem. □

5.1 Blocking SSTs of Diameter at Most 3 Is Insufficient

The example presented in Fig. 9 shows that blocking SSTs of diameter at most 3 is not sufficient even for blocking SSTs of diameter 4. In the example, it is clear that T (whose edges are represented by dotted lines) is an SST of diameter 4 and that the path B (whose edges are represented by full lines) avoids it. It is also clear that B blocks any SST of diameter 2, since such SSTs are stars, and B is a spanning subgraph of G. In order to prove that B meets all SSTs of diameter 3, we show that no edge in $E(G)$ can be the *central edge* of an SST of diameter 3 that avoids B. We do this using the following observation.

Lemma 5.3. *Suppose $[x,y] \in E(G) \setminus E(B)$. If there exist $z,w \in V(G)$ such that*

1. *The points x,y,z,w are distinct and in convex position,*
2. *$[x,w],[y,z] \in E(B)$, and*
3. *The segments $[x,y]$ and $[z,w]$ do not cross, then $[x,y]$ cannot be the central edge in an SST of diameter 3 that avoids B.*

Proof. Assume, on the contrary, that $[x,y]$ is the central edge of such an SST T. Then z and w must be at distance 1 in T from $[x,y]$. As $[x,w],[y,z] \in E(B)$ and T avoids B, this can happen only if $[x,z],[y,w] \in E(T)$. However, since x,y,z,w are in convex position and the pairs of edges $\{[x,y],[z,w]\}$ and $\{[x,w],[y,z]\}$ do not cross, the pair $\{[x,z],[y,w]\}$ must cross, contradicting the assumption that T is simple. □

It can be seen, by checking all pairs (i,j) with $1 \le i < j \le 7$, that in the example, no edge $[a_i,a_j]$ can be the central edge of an SST T of diameter 3 that avoids B, since for any edge $[a_i,a_j]$, at least one of the following holds:

1. $[a_i,a_j] \in E(B)$. This happens when $j = i+1$.
2. There exists a k such that $[a_i,a_k],[a_j,a_k] \in E(B)$ (and thus, a_k cannot be at distance 1 in T from $[a_i,a_j]$). This happens when $j = i+2$.
3. The vertices $x = a_i, y = a_j, z = a_{j+1}$, and $w = a_{i-1}$ satisfy the conditions of Lemma 5.3. This happens when $1 < i$ and $i+3 \le j < 7$. Note that in this case we never obtain $\{3,4,5\} \subset \{i-1,i,j,j+1\}$.
4. The vertices $x = a_i, y = a_j, z = a_{j-1}$, and $w = a_{i+1}$ satisfy the conditions of Lemma 5.3. This happens when $1 \le i < i+3 \le j \le 7$, and $i = 1$ or $j = 7$ (or both).

Therefore, B blocks all SSTs of diameter 3, as asserted.
We note that the example can be enlarged arbitrarily: The edges $[a_1,a_2]$ and $[a_6,a_7]$ can be replaced by longer convex polygonal arcs.

6 The Converse Direction

In this section, we prove Theorem 1.5, which is an improved variant of the converse direction of Theorem 1.2.

Theorem 6.1. *Let G be a complete geometric graph, and let $B \subset G$ be a comb in G. Then B meets every simple spanning subgraph of G.*

Proof. Assume, on the contrary, that H is a simple spanning subgraph of G such that $E(H) \cap E(B) = \emptyset$. Denote the spine of B by $\langle a_{-1},a_0,a_1,\ldots,a_k,a_{k+1}\rangle$. First, we show that there is no loss of generality in assuming that H does not contain edges of the form $[a_i,a_j]$, where $-1 \le i,j \le k+1$ (except, possibly, $[a_{-1},a_{k+1}]$).

Assume that H contains such edges, and let $[a_{i_0},a_{j_0}] \in H$, where $i_0 < j_0$, be such an edge that minimizes the difference $j-i$. Consider the subgraph of G defined by

$$G_1 = G \cap \mathrm{conv}(\{a_{i_0},a_{i_0+1},\ldots,a_{j_0}\}).$$

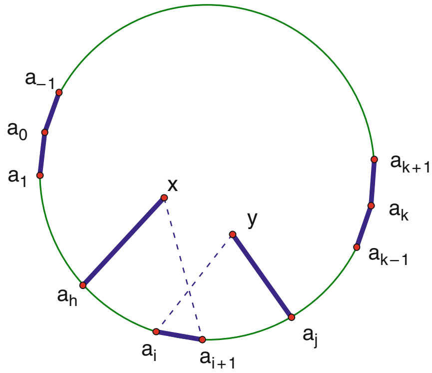


Fig. 10 An illustration to the proof of Theorem 6.1. The edges of B are represented by *full lines*, and the edges of H are represented by *dotted lines*

Note that H does not contain edges that connect vertices in $V(G_1)\setminus\{a_{i_0},a_{j_0}\}$ with vertices in $V(G)\setminus V(G_1)$, as such edges would cross the edge $[a_{i_0},a_{j_0}]$, contradicting the simplicity of H. Thus, the restriction of H to conv($\{a_{i_0},a_{i_0+1},\dots,a_{j_0}\}$), i.e., $H_1 = H\cap G_1$, is a simple spanning subgraph of G_1. Similarly, it is clear that $B_1 = B\cap G_1$ is a simple spanning caterpillar of G_1 such that the line spanned by an edge e of B_1 never crosses another edge of B_1. Finally, H_1 does not contain edges of the form $[a_i,a_j]$ for $i_0 \le i, j \le j_0$ (except for $[a_{i_0},a_{j_0}]$) by the minimality of the edge $[a_{i_0},a_{j_0}]$. Thus, by reduction to G_1, B_1, and H_1, we can assume w.l.o.g. that H does not contain edges of the form $[a_i,a_j]$ (except, possibly, for $[a_{-1},a_{k+1}]$).

Now, we define a function $f : \{0,1,\dots,k\} \to \{0,1,\dots,k\}$ by the following procedure, performed for each $0 \le i \le k$.

1. Consider the vertex a_i. Since H is a spanning subgraph, there exist edges in $E(H)$ that emanate from a_i. Pick one such edge $[a_i,y]$. Note that $y \ne a_l$ for $-1 \le l \le k+1$, since by assumption, H does not contain edges of the form $[a_i,a_l]$.
2. Since $y \ne a_l$ for all $-1 \le l \le k+1$, y is a leaf of B. Hence, the only edge of B that emanates from y connects it to an interior vertex of the spine of B, i.e., is of the form $[y,a_j]$ for some $0 \le j \le k$.
3. Define $f(i) = j$, where j is determined by the two previous steps.

Note that we indeed have $0 \le f(i) \le k$ for all i, since a_{-1} and a_{k+1} are leaves of B, and thus are connected only to a_0 and a_k, respectively. Also, note that $f(i) \ne i$ for all i, since otherwise B and H would share the edge $[a_i,y]$, for some y.

Consequently, we have $f(0) > 0$ and $f(k) < k$, and thus, there exists an i, $0 \le i \le k$, such that $j = f(i) > i$ and $h = f(i+1) < i+1$. Denote the vertices that were used in the generation of $f(i)$ and of $f(i+1)$ by y and x, respectively, as illustrated in Fig. 10. We claim that the edges $[a_i,y]$ and $[a_{i+1},x]$ cross, which contradicts the assumption that H is simple.

In order to prove this claim, consider the following polygon:

$$P = \langle x, a_h, a_{h+1}, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{j-1}, a_j, y, x\rangle.$$

(Note that the highest possible value of h is i. If $h = i$, then there are no edges between a_h and a_i; similarly for a_{i+1} and a_j.) We claim that the path $P_0 = \langle x, a_h, \ldots, a_i, a_{i+1}, \ldots, a_j, y \rangle$ lies on the boundary of conv(P). Indeed, all the edges of P_0 except for $[x, a_h]$ and $[a_j, y]$ lie on the boundary of conv(G), so they clearly support P. The edge $[a_h, x]$ also supports P, since the line $l(a_h, x)$ meets the path $\langle a_h, \ldots, a_i, a_{i+1}, \ldots, a_j, y \rangle$ only at its endpoint a_h. For similar reasons, $[a_j, y]$ must also support P. It follows that the path P_0 is part of the boundary of conv(P).

Finally, a_{i+1} lies on the boundary of conv(P) strictly between a_i and y, and a_i lies on the boundary of conv(P) strictly between a_{i+1} and x. This implies that the two edges $[a_i, y]$ and $[a_{i+1}, x]$ must cross. This contradicts the assumption that H is simple, and thus completes the proof of the theorem. □

References

1. F. Harary, A.J. Schwenk, Trees with Hamiltonian square. Mathematika **18**, 138–140 (1971)
2. F. Harary, A.J. Schwenk, The number of caterpillars. Disc. Math. **6**, 359–365 (1973)
3. M. Carmen Hernando, Complejidad de Estructuras Geométricas y Combinatorias, Ph.D. Thesis, Universitat Politéctnica de Catalunya, 1999 [in Spanish]. Available online at: http://www.tdx.cat/TDX-0402108-120036/
4. G. Károlyi, J. Pach, G. Tóth, Ramsey-type results for geometric graphs I. Discrete Comput. Geom. **18**, 247–255 (1997)
5. C. Keller, M.A. Perles, On the smallest sets blocking simple perfect matchings in a convex geometric graph. Israel J. Math. **187**, 465–484 (2012)
6. C. Keller, M.A. Perles, Characterization of co-blockers for simple perfect matchings in a convex geometric graph, submitted. Available online at: http://arxiv.org/abs/1011.5883.
7. J.J. Montellano-Ballesteros, E. Rivera-Campo, On the heterochromatic number of hypergraphs associated to geometric graphs and to matroids, to appear in Graphs and Combinatorics. DOI: 10.1007/s00373-012-1190-y
8. J. Pach, in *Geometric Graph Theory*, ed. by J.E. Goodman, J. O'Rourke. Handbook of Discrete and Computational Geometry, 2nd edn. Chapter 10 (CRC Press, Boca Raton, FL, 2004), pp. 219–238

Coloring Clean and K_4-Free Circle Graphs

Alexandr V. Kostochka and Kevin G. Milans

Abstract A *circle graph* is the intersection graph of chords drawn in a circle. The best-known general upper bound on the chromatic number of circle graphs with clique number k is $50 \cdot 2^k$. We prove a stronger bound of $2k-1$ for graphs in a simpler subclass of circle graphs, called *clean graphs*. Based on this result, we prove that the chromatic number of every circle graph with clique number at most 3 is at most 38.

1 Introduction

Recall that the *chromatic number* of a graph G, denoted $\chi(G)$, is the minimum size of a partition of $V(G)$ into independent sets. A *clique* is a set of pairwise adjacent vertices, and the *clique number* of G, denoted $\omega(G)$, is the maximum size of a clique in G.

Vertices in a clique must receive distinct colors, so $\chi(G) \geq \omega(G)$ for every graph G. In general, $\chi(G)$ cannot be bounded above by any function of $\omega(G)$. Indeed, there are triangle-free graphs with arbitrarily large chromatic number [4, 18].

When graphs have additional structure, it may be possible to bound the chromatic number in terms of the clique number. A family of graphs $\mathcal{G}$ is χ*-bounded* if there is a function f such that $\chi(G) \leq f(\omega(G))$ for each $G \in \mathcal{G}$. Some families of intersection graphs of geometric objects have been shown to be χ-bounded (see e.g. [8, 11, 12]). Recall that the *intersection graph* of a family of sets has a vertex

A.V. Kostochka (✉)
Department of Mathematics, University of Illinois at Urbana-Champaign and Sobolev Institute of Mathematics, Novosibirsk, Russia
e-mail: kostochk@math.uiuc.edu

K.G. Milans
Department of Mathematics, University of South Carolina, Columbia, South Carolina, USA
e-mail: milans@math.sc.edu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_21,

for each set in the family, with vertices adjacent if and only if the corresponding sets intersect. Possibly the simplest example is the class $\mathcal{I}$ of *interval graphs*, i.e., the class of intersection graphs of intervals in a line. Interval graphs are perfect graphs; i.e., $\chi(G) = \omega(G)$ for every interval graph G. Another interesting family is the family $\mathcal{C}$ of *circle graphs*, that is, the intersection graphs of families of chords of a circle. This family is more complicated than $\mathcal{I}$: Although the problem of recognition of a circle graph is polynomial (Bouchet [2]) and so are the problems of finding maximum cliques and maximum independent sets in circle graphs (Gavril [7, 8]), the problems of finding the chromatic number (Garey et al. [6]) and clique covering number (Keil and Stewart [14]) are NP-complete. Circle graphs naturally arise in a number of combinatorial problems: from sorting problems to studying planar graphs to continuous fractions (see, e.g., [3, 5, 8]).

A graph G is a circle graph if and only if it is an *overlap graph*: The vertices of such a graph are closed intervals in the real line, and two intervals are adjacent if they *overlap*, that is, intersect and neither of them contains the other. To see this, observe that given a family of chords on a circle representing a circle graph, cutting the circle at a point and unrolling gives an overlap representation for the same graph.

The above-mentioned complexity results on circle graphs make interesting upper bounds on the chromatic number of circle graphs in terms of their clique number, especially if the proofs yield polynomial-time algorithms for corresponding colorings. There was a series of results in this direction. Karapetyan [13] showed that $\chi(G) \leq 8$ when G is a triangle-free circle graph. Gyárfás [9, 10] proved that $\chi(G) \leq k^2 2^k(2^k - 2)$ when G is a circle graph with clique number k. The bound was improved in [17] to $\chi(G) \leq k(k+2)2^k$, and in [16] to $\chi(G) \leq 50 \cdot 2^k - 32k - 64$. The best-known lower bound for the maximum chromatic number of circle graphs with clique number k is only $0.5k(\ln k - 2)$ [15, 17]. The exponential gap has remained open for 25 years.

Exact results are known only for circle graphs with clique number at most 2. Kostochka [17] showed that $\chi(G) \leq 5$ for every such graph G, and Ageev [1] constructed a triangle-free circle graph with chromatic number 5.

The purpose of this chapter is twofold. First, we consider a simple subclass of circle graphs, the *clean graphs*. A family of intervals X is *clean* if no interval is contained in the intersection of two overlapping intervals in X. A circle graph is *clean* if it is the overlap graph of a clean family of intervals. Since the structure of clean graphs is much simpler than that of general circle graphs, we are able to prove a much better bound for clean graphs.

Theorem 1.1. *For every clean circle graph G with clique number k, $\chi(G) \leq 2k - 1$.*

Moreover, the proof yields a polynomial-time algorithm that, for each clean circle graph G with clique number k, finds a $(2k-1)$-coloring of a special type, a *good coloring* that will be defined later. On the other hand, we show that for every k, there exists a clean circle graph G with clique number k that needs $2k - 1$ colors for a good coloring.

We use Theorem 1.1 to derive an upper bound on the chromatic number of K_4-free circle graphs. For such graphs G, the general bound in [17] implies that $\chi(G) \leq 120$. Our second main results follows.

Theorem 1.2. *For every circle graph G with clique number at most* 3, $\chi(G) \leq 38$.

It can be checked that the proof of Theorem 1.2 yields a polynomial-time algorithm for coloring K_4-free circle graphs using at most 38 colors. In the next section, we introduce notation and basic concepts. In Sect. 3, we prove Theorem 1.1, and in the last section, we prove Theorem 1.2.

2 Preliminaries

Recall that a circle graph is the intersection graph of a set of chords of a circle. Before proceeding, we clarify a minor technical point. Suppose a set of chords share a common endpoint p on the circle but are otherwise internally disjoint. Should these chords be considered pairwise intersecting or pairwise disjoint? Fortunately, the answer does not matter: Without affecting other intersection relationships, one can spread the common endpoints over a small portion of the circle near p so that the chords either intersect or are disjoint, as desired. Therefore, we may (and will) assume that the endpoints of all chords are distinct.

By the discussion in the previous section, for every circle graph F, there exists an *overlap representation* of F, that is, a family X of intervals in the real line such that F is the overlap graph of X. In this case, we also will write that $F = G(X)$. Since the endpoints of chords in a circle representation are distinct, the intervals in X have distinct endpoints.

Definition 2.1. An interval $[a,b]$ is a *left-neighbor* of $[c,d]$ if $a < c < b < d$. We use $L_X(u)$ to denote the set of all left-neighbors of an interval u in a family X, or simply $L(u)$ when X is clear from the context. Similarly, $[a,b]$ is a *right-neighbor* of $[c,d]$ if $c < a < d < b$, and $R_X(u)$ denotes the set of all right-neighbors of u. We also define the *closed* left- and right-neighborhoods via $\overline{L}_X(u) = L_X(u) \cup \{u\}$ and $\overline{R}_X(u) = R_X(u) \cup \{u\}$. For each interval u, we use $l(u)$ to denote the left endpoint of u and $r(u)$ to denote the right endpoint of u.

The *inclusion order* is defined by containment. The *endpoint order* is defined by putting $x \leq y$ if and only if $l(x) \leq l(y)$ and $r(x) \leq r(y)$. Note that $x \leq y$ in the endpoint order if and only if x comes before y in both the left-endpoint order and the right-endpoint order. Note that any two distinct intervals are comparable in exactly one of the inclusion order and the endpoint order.

Definition 2.2. If S is a set of intervals, then the *center* of S is the intersection of the intervals in S. A family of intervals X is *clean* if no interval is contained in the intersection of two overlapping intervals. A circle graph is *clean* if it is the overlap graph of a clean family of intervals.

A set S of vertices in a graph G is a *cutset* if $G - S$ is disconnected. When S is a cutset in G, the graphs induced by the union of S and the vertices of a component of $G - S$ are *S-lobes*. To color G, it suffices to color the S-lobes so that the colorings agree on S. To ensure that S is colored in the same way in all S-lobes, our inductive hypothesis prescribes the way in which S is colored.

Definition 2.3. A subset A of a poset P is a *downset* if $y \in A$ whenever $y \leq x$ for some $x \in A$, and A is an *upset* if $y \in A$ whenever $y \geq x$ for some $x \in A$. For an element $z \in P$, we use $D[z]$ to denote the downset $\{y \in P\colon\ y \leq z\}$ and $D(z)$ to denote the downset $\{y \in P\colon\ y < z\}$. Similarly, $U[z]$ denotes the upset $\{y \in P\colon\ y \geq z\}$ and $U(z)$ denotes the upset $\{y \in P\colon\ y > z\}$. The *height* of an element $x \in X$ is the size of a maximum chain in $D[x]$ and the *depth* of x is the size of a maximum chain in $U[x]$. When X is a family of intervals, we define $h_X(x)$ [or simply $h(x)$ when X is clear from the context] to be the height of x in the endpoint order on X. The *canonical coloring* of a family X of intervals assigns $h(x)$ to each interval $x \in X$. A coloring f of a family X of intervals is *canonical*, and we say that f is *canonical* on X if the color classes of f form the same partition of X as the color classes of the canonical coloring.

Note that the canonical coloring is a proper coloring; if x and y overlap, then they are comparable in the endpoint order, and therefore $h(x) \neq h(y)$.

3 Clean Circle Graphs

Definition 3.1. A coloring f of a family of intervals X is *good* if, for each $w \in X$, f is canonical on $\overline{R}(w)$.

Note that if f is a good coloring of X, then it follows that f is a proper coloring. While some families of intervals do not admit good colorings with any number of colors, clean families have good colorings. The goal of this section is to prove the following refinement of Theorem 1.1.

Theorem 3.2. *If X is a clean family of intervals with clique number $k \geq 1$, then there is a good coloring f of $G(X)$ using at most $2k-1$ colors.*

Proposition 3.3. *In a clean family of intervals, let x be an interval with $h(x) \geq 2$. If y is chosen from $D(x)$ to maximize $l(y)$, then $h(x) = h(y) + 1$.*

Proof. Let $k = h(x)$; we use induction on k. When $k = 2$, the statement is trivial. Suppose $k \geq 3$. Since $h(y) < h(x)$, it suffices to show that $h(y) \geq h(x) - 1$. Since $h(x) = k$, there is a chain $z_1, \ldots, z_k$ with $z_k = x$. We may assume that $y \neq z_{k-1}$, so the choice of y yields $l(z_{k-1}) < l(y)$. Therefore, $l(z_{k-2}) < l(z_{k-1}) < l(y)$. Consider the order of $r(y)$ and $r(z_{k-2})$. If $r(y) < r(z_{k-2})$, then y is contained in $z_{k-1} \cap z_{k-2}$, contradicting that the family is clean. Otherwise, $r(y) > r(z_{k-2})$; now $y > z_{k-2}$ and $h(y) \geq k-1$. □

Proposition 3.4. *Let X be a clean family, and let x be an interval in X that contains another interval in X. If $Y = X - \{x\}$, then $h_Y(u) = h_X(u)$ for all $u \in Y$.*

Proof. Let z be an interval in X that is contained in x. Because X is clean, $y < x$ implies $y < z$, and $y > x$ implies $y > z$. Therefore, if C is a chain containing x, then substituting z for x in C yields another chain of the same size. Hence, $h_Y(u) \geq h_X(u)$ for all $u \in Y$, and the other inequality holds since $Y \subseteq X$. □

Proposition 3.5. *If X is a family of intervals that share a common point a and the overlap graph $G(X)$ has clique number k, then the canonical coloring on X uses exactly k colors.*

Proof. Because the canonical coloring is proper, it uses at least k colors. For the other direction, if the canonical coloring uses r colors, then there is a chain C of size r in the endpoint order on X. It follows that C is an independent set in the inclusion order on X. Hence, no two intervals in C are related by containment. However, C is pairwise intersecting because every member of X contains a. It follows that the intervals in C pairwise overlap, and so $k \geq r$. □

Proposition 3.6. *If f is canonical on X, and Y is a downset of X in the endpoint order, then f is canonical on Y.*

Proof. Because Y is a downset in X, we have $h_X(x) = h_Y(x)$ for each $x \in Y$. □

Let X be a family of intervals and let $u \in X$ be an interval that is not inclusion-minimal, where $u = [a,b]$. We define the *subordinate* of u to be the interval with the rightmost right endpoint among all intervals contained in u. Let v be the subordinate of u, where $v = [c,d]$, and define the *modified subordinate* to be the interval v', where $v' = [c,b]$. The *right-push* operation on u produces the families Y and Y' and a map $\phi : Y \to Y'$, where $Y = X - u$, $Y' = Y - v + v'$, and $\phi(x) = x$ for $x \neq v$ and $\phi(v) = v'$. When $S \subseteq X$, we write $\phi(S)$ for $\{\phi(w) : w \in S\}$.

Lemma 3.7. *Let X be a family of intervals, let $u \in X$ be an interval that is not inclusion-minimal, and let Y, Y', and $\phi : Y \to Y'$ be produced by the right-push operation on u. If X is clean, then the following hold:*

(1) The map ϕ preserves the order of the left and right endpoints. That is, $l(x) < l(y)$ if and only if $l(\phi(x)) < l(\phi(y))$ for each $x, y \in Y$. Similarly, $r(x) < r(y)$ if and only if $r(\phi(x)) < r(\phi(y))$.
(2) Y and Y' are clean.
(3) The clique numbers of Y and Y' are both at most the clique number of X.
(4) For each $w \in Y$ with $w \neq v$, we have $\phi(\overline{R}_Y(w)) = \overline{R}_{Y'}(\phi(w))$.
(5) $\phi(\overline{R}_Y(v)) \subseteq \overline{R}_{Y'}(\phi(v))$ and $\phi(\overline{R}_Y(v))$ is a downset of $\overline{R}_{Y'}(\phi(v))$ in the endpoint order.

Proof. Define a,b,c,d so that $u = [a,b]$ and $v = [c,d]$.

(1) Note that no interval in X has its right endpoint strictly between d and b. Indeed, if there were such an interval w, then either w is contained in u, in which case v is not the subordinate of w, or $\{w,u\}$ is a 2-clique whose center contains v,

contradicting that X is clean. In passing from Y to Y', the right endpoint of v is moved from d to b to form a new interval v'. Doing so preserves the order of the left and right endpoints.

(2) Because $Y \subseteq X$ and X is clean, we have that Y is clean. Note that x is contained in the center of a 2-clique with $\{y,z\}$ with $y \le z$ if and only if $l(y) < l(z) < l(x)$ and $r(x) < r(y) < r(z)$. Hence, the property of being clean is determined by the order of the left endpoints and the order of right endpoints. Because ϕ preserves these orders, Y' is also clean.

(3) Let k be the clique number of X. Because $Y \subseteq X$, the clique number of Y is at most k. Let $\{x_1,\ldots,x_t\}$ be a clique S in Y' with $x_1 < \cdots < x_t$, and note that $l(x_1) < \cdots < l(x_t) < r(x_1) < \cdots < r(x_t)$. Suppose for a contradiction that $t > k$. We have that $x_j = v'$ for some j, or else S is a clique in Y. If $j > 1$, then x_{j-1} cannot have its right endpoint between d and b. Because $d < b$ and $r(x_j) = b$, it follows that $r(x_{j-1}) < d < r(x_j)$. But $d = r(v)$, and so $S - v' + v$ is a clique of size t in Y, a contradiction. Hence, it must be that $j = 1$. Recalling that $r(x_1) = r(u) = b$, we have that $l(u) < l(x_2) < \cdots < l(x_t) < r(u) < r(x_2) < \cdots < r(x_t)$, which implies that $S - v' + u$ is a clique of size t in X, another contradiction.

(4) If $x \in \overline{R}_Y(w)$, then passing from x to $\phi(x)$ leaves the left endpoint fixed and possibly increases the right endpoint. Because $w \neq v$ and $\phi(w) = w$, it follows that $\phi(x) \in \overline{R}_{Y'}(\phi(w))$ and so $\phi(\overline{R}_Y(w)) \subseteq \overline{R}_{Y'}(\phi(w))$. Conversely, if $\phi(x) \in \overline{R}_{Y'}(\phi(w))$, then passing from $\phi(x)$ to x leaves the left endpoint fixed and possibly decreases the right endpoint. However, the right endpoint of $\phi(x)$ must remain greater than the right endpoint of $\phi(w)$, and so $x \in \overline{R}_Y(w)$. It follows that $\overline{R}_{Y'}(\phi(w)) \subseteq \phi(\overline{R}_Y(w))$.

(5) Passing from v to v' increases the right endpoint of v, but in doing so, the right endpoint never crosses the right endpoint of another interval. Hence, each right-neighbor of v in Y is a right-neighbor of v' in Y', and therefore $\phi(\overline{R}_Y(v)) \subseteq \overline{R}_{Y'}(\phi(v))$. To see that $\phi(\overline{R}_Y(v))$ is a downset of $\overline{R}_{Y'}(v')$, note that a right-neighbor x of v' in Y' is also a right-neighbor of v in Y if and only if $l(x) < d$. □

Note that because the endpoint order on X only depends on the order of the left endpoints and that of the right endpoints, a consequence of Lemma 3.7 is that ϕ is a poset isomorphism from Y to Y' under the endpoint order.

Proposition 3.8. *Let X be a clean family of intervals and let $u \in X$ be a nonminimal element in the inclusion order. If v is chosen from $\{w \in X : w \subseteq u\}$ to minimize the left endpoint, then $h_X(u) = h_X(v)$.*

Proof. We argue that $w < u$ if and only if $w < v$. If $w < u$, then also $w < v$ or else $\{w,u\}$ is a 2-clique with v in the center, contradicting that X is clean. Conversely, if $w < v$, then the extremality of v implies that $w < u$. □

Lemma 3.9. *If X is a clean family and f is the canonical coloring on X, then f is good.*

Proof. Let $z \in X$, let $S = \overline{R}_X(z)$, and let h_S (resp., h_X) be the height function for the endpoint order on S (resp., X). We show that $h_S(u) = h_S(v)$ if and only if $h_X(u) = h_X(v)$ for each $u, v \in S$. For each $k \geq 1$, let $T_k = \{w \in S : h_S(w) = k\}$. Fix a positive integer k. Because all elements in T_k have the same height, they are not comparable in the endpoint order, and therefore T_k is a chain in the inclusion order. Index the elements of T_k as $u_1, \ldots, u_n$ so that $u_1 \subsetneq u_2 \subsetneq \cdots \subsetneq u_n$, and fix $j < n$. We claim there are no intervals in X whose left endpoint is between $l(u_j)$ and $l(u_{j+1})$. Indeed, if there are such intervals, then let v be one that minimizes the left endpoint. Note that $v \subsetneq u_{j+1}$, or else $\{u_{j+1}, v\}$ is a 2-clique with u_j in the center. Also, $v \notin S$, or else applying Proposition 3.8 to u_{j+1} and v in the family S would give that $h_S(v) = h_S(u_{j+1}) = k$, and hence $v \in T_k$, a contradiction because no interval in T_k has a left endpoint between a left endpoints of u_j and u_{j+1}. But now $v \notin S$ implies that v is in the center of the 2-clique $\{z, u_{j+1}\}$, a contradiction. A final application of Proposition 3.8 to u_{j+1} and u_j in X gives that $h_X(u_{j+1}) = h_X(u_j)$. It follows that all intervals in T_k have the same height in X.

For the converse, suppose that T_k and $T_{k'}$ with $k < k'$ have the property that all elements in $T_k \cup T_{k'}$ have the same height in the endpoint order on X. Fix $u \in T_{k'}$. Because $h_S(u) = k'$ and $k < k'$, there is an interval $v \in S$ with $v < u$ and $h_S(v) = k$. It follows that $v \in T_k$. But now v and u are comparable in the endpoint order, so they cannot have the same height in X. □

Lemma 3.10. *Let X be a clean family of intervals, let u be a nonminimal element in the inclusion order on X, and obtain Y, Y', and ϕ from the right-push operation on u. If g' is a good coloring of Y', then $g' \circ \phi$ is a good coloring of Y.*

Proof. Consider $w \in Y$. Because g' is good on Y', we have that g' is canonical on $\overline{R}_{Y'}(\phi(w))$. By Lemma 3.7, we have that $\phi(\overline{R}_Y(w))$ is a downset of $\overline{R}_{Y'}(\phi(w))$ in the endpoint order (even equality holds when $w \neq v$). By Proposition 3.6, we have that g' is canonical on $\phi(\overline{R}_Y(w))$. But $\phi : Y \to Y'$ is an isomorphism of the endpoint orders on Y and Y', so $g' \circ \phi$ is canonical on $\overline{R}_Y(w)$. □

If $a \in \mathbb{R}$, then X^a denotes the subfamily of X consisting of all intervals that contain a in their interior, $X^{>a}$ denotes $\{[c,d] \in X\colon\ c > a\}$, and $X^{<a}$ denotes $\{[c,d] \in X\colon\ d < a\}$.

Proposition 3.11. *Let f be a good coloring of X, let α and β be colors, let a be a point on the real line, and suppose that $f(u) \notin \{\alpha, \beta\}$ for each $u \in X^a$. If f' is the coloring of X obtained from f by interchanging α and β on the intervals in $X^{>a}$, then f' is also good.*

Proof. Let $w \in X$ and define c, d so that $w = [c,d]$. If $d > a$, then every interval in $\overline{R}_X(w)$ with a color in $\{\alpha, \beta\}$ is in $X^{>a}$, and so the change in colors does not alter the partition on $\overline{R}_X(w)$ given by the color classes of f. Similarly, if $d < a$, then every interval in $\overline{R}_X(w)$ with a color in $\{\alpha, \beta\}$ is in $X^{<a}$, and so none of these intervals changes colors. If $d = a$, then increase a by a small amount and apply the proposition again. □

We are now ready to prove Theorem 3.2.

Proof. By induction on $|X|$; we may assume $|X| \geq 1$ and $k \geq 2$. Let x be the interval in X that minimizes $l(x)$. If $R(x) = \emptyset$, then x has no neighbors. Let $Y = X - x$, apply induction to Y to obtain a good coloring g of Y, and extend g to a coloring f of X by assigning an arbitrarily chosen color to x. Clearly, f is canonical on each right-neighborhood.

Therefore, we may assume that x has right-neighbors. Choose $y \in R(x)$ to minimize $l(y)$, and define a and b so that $y = [a,b]$. Let $Y_1 = \{z \in X\colon\ l(z) \leq b\}$ and $Y_2 = \{z \in X\colon\ r(z) \geq b\}$. Note that $x \notin Y_2$ and therefore $Y_2 \subsetneq X$. If also $Y_1 \subsetneq X$, then we may apply induction to Y_1 and Y_2 to obtain respective good colorings g_1 and g_2. Note that $Y_1 \cap Y_2 = \{z \in X\colon\ l(z) \leq b \leq r(z)\}$, and because y is inclusion-maximal, $Y_1 \cap Y_2 = \overline{R}_X(y)$. Consequently, all right-neighbors of y survive in Y_1 and Y_2, and hence $\overline{R}_X(y) = \overline{R}_{Y_1}(y) = \overline{R}_{Y_2}(y)$, which implies that g_1 and g_2 are canonical on $Y_1 \cap Y_2$. Hence, after relabeling the color names, we obtain a coloring g of X that is a common extension of g_1 and g_2. Clearly, g uses at most $2k-1$ colors; it remains to show that g is canonical on each right-neighborhood. Consider $u \in X$. If $r(u) \leq b$, then $\overline{R}_X(u) \subseteq Y_1$ and so $\overline{R}_X(u) = \overline{R}_{Y_1}(u)$, which implies that g is canonical on $\overline{R}_X(u)$. Otherwise, $\overline{R}_X(u) \subseteq Y_2$, and so $\overline{R}_X(u) = \overline{R}_{Y_2}(u)$, which again implies that g is canonical on $\overline{R}_X(u)$.

Hence, we may assume $X = Y_1$. Next, we consider the case that x is not inclusion-minimal. Let v be the subordinate of x, let v' be the modified subordinate of x, and obtain Y, Y', and ϕ from the right-push operation on x. By Lemma 3.7, we have that Y and Y' are clean with clique number at most k. By induction and Lemma 3.10, we obtain good colorings g' of Y' and $g_0 = g' \circ \phi$ of Y using at most $2k-1$ colors. Extend g_0 to a coloring g of X by defining $g(w) = g_0(w)$ for $w \neq x$ and $g(x) = g_0(v) = g'(v')$. Clearly, g uses at most $2k-1$ colors. We claim that g is a good coloring. First, note that because x minimizes $l(x)$, we have that $x \in \overline{R}_X(w)$ implies that $w = x$. Therefore, g inherits the canonical coloring of g_0 on $\overline{R}_X(w)$ whenever $w \neq x$. Finally, note that because X is clean, we have that $R_X(x) = R_{Y'}(v')$ and hence g inherits the canonical coloring on $\overline{R}_X(x)$ from the canonical coloring of g' on $\overline{R}_{Y'}(v')$.

Hence, we may assume that x is inclusion-minimal; it follows that $y \in \overline{R}_X(w)$ implies that $w \in \{x,y\}$. Next, we consider the case that y is not inclusion-minimal. Let v be the subordinate of y, let v' be the modified subordinate, and obtain Y, Y' and ϕ from the push operation. By Lemma 3.7, we have that Y and Y' are clean with clique number at most k. By induction and Lemma 3.10, obtain good colorings g' of Y' and $g_0 = g' \circ \phi$ of Y using at most $2k-1$ colors. We use g_0 to construct a good coloring of X. Because $Y = X - x$, to extend a good coloring of Y to a good coloring of X, we must assign a color to y so that the coloring remains canonical on each closed right-neighborhood. Because y is only in the closed right-neighborhood of x and y, we need only verify that the coloring is canonical on $\overline{R}_X(x)$ and $\overline{R}_X(y)$.

We consider two subcases. First, suppose that y is inclusion-minimal in $\overline{R}_X(x)$. Because y is chosen from $\overline{R}_X(x)$ to minimize $l(y)$, it follows that $x < y < z$ for every $z \in \overline{R}_X(x) - \{x,y\}$. With $Z_1 = \overline{R}_X(x)$ and $Z_2 = \overline{R}_Y(x) = \overline{R}_X(x) - \{y\}$, this implies that two elements have the same height in Z_2 if and only if they have the same height in Z_1, and y is the only element of height 1 in Z_1. Consequently, an extension of g_0 to X is canonical on $\overline{R}_X(x)$ if and only if it assigns y a color that is not used on any other interval in $\overline{R}_X(x)$. Similarly, $y < z$ for each $z \in \overline{R}_X(y) - \{y\}$ and hence an extension of g_0 to X is canonical on $\overline{R}_X(y)$ if and only if y is assigned a color that is not used on any other interval in $\overline{R}_X(y)$. Because g_0 is canonical on $\overline{R}_Y(x)$ and the clique number of $\overline{R}_Y(x)$ is at most $k-1$ [indeed, every maximal clique in $\overline{R}_X(x)$ contains y], it follows that g_0 uses at most $k-1$ colors on $\overline{R}_Y(x)$. Also, g' uses at most k colors on $\overline{R}_{Y'}(v')$, and hence g_0 uses at most $k-1$ colors on $R_X(y)$ [indeed, $g'(v')$ is used on $v' \in \overline{R}_{Y'}(v')$ but is not used on any interval in $R_X(y)$]. Because $2k-1$ colors are available and at most $2k-2$ provide conflicts, one color remains available for assignment to y.

The second subcase is that y is not inclusion-minimal in $\overline{R}_X(x)$. Let z be the interval that minimizes $l(z)$ among all intervals in $\overline{R}_X(x)$ that are contained in y. Note that z is also the interval that minimizes $l(z)$ among all that are contained in y. Let $\alpha = g_0(z)$. By Proposition 3.8, the height of y and the height of z are the same in all subsets of X containing z and y. By induction, we have that g_0 is canonical on $\overline{R}_X(x) - y$. Applying Proposition 3.4 to $\overline{R}_X(x)$, an extension of g_0 to X is canonical on $\overline{R}_X(x)$ if and only if y is assigned color α. Also, an extension of g_0 to X is canonical on $\overline{R}_X(y)$ if and only if y is assigned a color different from every other interval in $\overline{R}_X(y)$. If α is not used on $R_X(y)$, then we may assign α to y. Otherwise, we first modify g_0 before extending to X. Note that z is inclusion-maximal in Y, and let a be a point slightly to the right of $r(z)$. Because z is inclusion-maximal in Y, every interval in Y that contains a is in $R_Y(z)$. Let A be the set of colors that g_0 uses on intervals containing a. Because g_0 is canonical on $\overline{R}_Y(z)$, at most k colors are used on these intervals; because g_0 uses α on $z \in \overline{R}_Y(z)$, we have $\alpha \notin A$ and hence $|A| \leq k-1$. Let B be the set of colors that g_0 uses on intervals in $R_X(y)$. Because g' is canonical on $\overline{R}_{Y'}(v')$, $R_X(y) = \overline{R}_{Y'}(v') - \{v'\}$, and v' overlaps with every other interval in $\overline{R}_{Y'}(v')$, we have that $|B| \leq k-1$. Let β be a color that g_0 uses but is not contained in $A \cup B$. Obtain g_1 from g_0 by applying Proposition 3.11 with colors $\{\alpha, \beta\}$ at point a. Note that because $\beta \notin B$, we have that g_1 does not use α on any interval in $R_X(y)$. Also, $g_1(z) = \alpha$ and an extension of g_1 to X is canonical on $\overline{R}_X(x)$ if and only if y is assigned color α. Therefore, we obtain a good coloring of X from g_1 by assigning y the color α.

Hence, we may assume that both x and y are inclusion-minimal. By Lemma 3.9, the canonical coloring on X is good. Because $X - x = \overline{R}_X(y)$ and Proposition 3.5 implies that the canonical coloring uses at most k colors on $\overline{R}_X(y)$, the canonical coloring on X uses at most $k+1$ colors in total. □

Theorem 3.12. *For each $k \geq 1$, there is a clean circle graph G with $\omega(G) = k$ such that every good coloring of G uses at least $2k-1$ colors.*

Proof. We construct G in stages. Our construction uses a set of $k-1$ intervals V that induce a clique in the overlap graph and a set of $k-1$ intervals V' that form a chain under inclusion. These intervals are represented by solid lines in Fig. 1, which presents the construction for $k = 5$. Let $V = \{v_1, \ldots, v_{k-1}\}$ and let

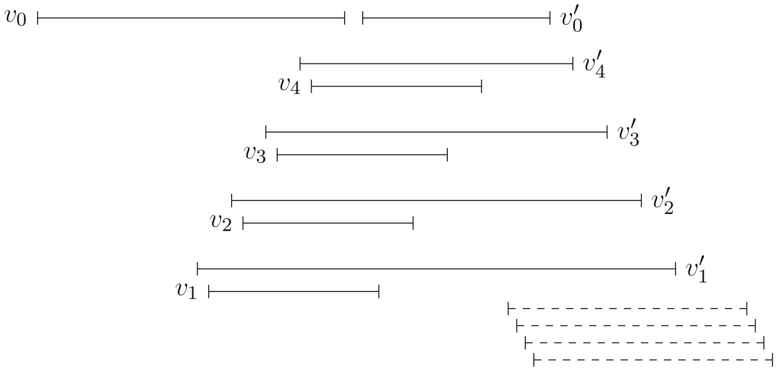


Fig. 1 Construction in Theorem 3.12

$V' = \{v_1', \ldots, v_{k-1}'\}$, indexed so that $v_1 < \cdots < v_{k-1}$ and $v_1' \supseteq \cdots \supseteq v_{k-1}'$. The left endpoint of v_j' is placed slightly to the left of $l(v_j)$, and the right endpoints of intervals in V' satisfy $r(v_1') \geq \cdots \geq r(v_{k-1}')$. Next, add v_0 so that v_0 is a left-neighbor of all intervals in $V \cup V'$, and add v_0' so that v_0' is a right-neighbor of all intervals in V and is contained in all intervals in V'.

Because a good coloring must be canonical on $\overline{R}(v_0)$, it follows that a good coloring assigns the same color to v_j and v_j' for $j \geq 1$, and hence $k-1$ distinct colors are assigned to intervals in V'. Since v_0' is a right-neighbor of each interval in V, it follows that k distinct colors are assigned to intervals in $V' \cup \{v_0'\}$. These intervals form an independent set in the overlap graph.

In the second stage, we add a set S of $k-1$ pairwise overlapping intervals such that each interval in S overlaps with intervals in $V' \cup \{v_0'\}$ and no others. Intervals in S are represented by dashed lines in Fig. 1. A good coloring must use $k-1$ new colors on S, and hence at least $2k-1$ colors in total. □

4 Chromatic Number of K_4-Free Circle Graphs

In this section, we study the chromatic number of circle graphs with clique number at most 3. By Theorem 3.2, it follows that a clean K_4-free circle graph has chromatic number at most 5. We need a lemma that provides 5-colorings of other circle graphs. Recall that if a is a point in $\mathbb{R}$, then X^a is the set of all intervals in X that contain a.

Lemma 4.1. *Let $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ be points with $a_1 < b_1 < a_2 < b_2 < \cdots < a_k < b_k$, and let $S_j = \{a_j, b_j\}$. Let X be a family of intervals, each of which has nonempty intersection with exactly one of the sets in $\{S_1, \ldots, S_k\}$. If $\omega(G(X)) \leq 3$ and $\omega(G(X^c)) \leq 2$ for each $c \in \{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_k\}$, then there is a proper 5-coloring of $G(X)$ with a distinguished color α such that every interval assigned color α is disjoint from $\{a_1, \ldots, a_k\}$, and for each $c \in \{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_k\}$, at most 4 colors are used on intervals in X^c.*

Proof. Partition X into $A_1, B_1, \ldots, A_k, B_k$ as follows. If S_j is the unique set in $\{S_1, \ldots, S_k\}$ that has a nonempty intersection with x, then we assign x to the set A_j if $a_j \in x$ or to the set B_j otherwise. Note that for all $1 \leq i, j \leq k$, we have $\omega(G(A_i)) \leq 2$ and $\omega(G(B_j)) \leq 2$, and hence the endpoint orders on A_i and B_j are posets of height at most 2.

The canonical coloring is defined with respect to height in the endpoint order. The *dual-canonical coloring* of a family of intervals colors each interval with its depth in the endpoint order. When the interval order on Z has height at most t, the $(\beta_1, \ldots, \beta_t)$-canonical coloring on Z assigns to an interval $z \in Z$ the color β_j, where j is the height of z in the endpoint order. Similarly, the $(\beta_1, \ldots, \beta_t)$-dual-canonical coloring on Z assigns to an interval $z \in Z$ the color β_j, where j is the depth of z in the endpoint order. We color each A_j canonically, and we color each B_j with a dual-canonical coloring.

We use $\{1, 2, 3, 4, \alpha\}$ as our set of colors. If j is odd, then we use the $(2, 1)$-canonical coloring on A_j and the $(\alpha, 3)$-dual-canonical coloring on B_j. If j is even, then we use the $(4, 3)$-canonical coloring on A_j and the $(\alpha, 1)$-dual-canonical coloring on B_j. First, note that if x has color α, then x is in some B_j, which implies that x contains b_j but not a_j, and therefore x does not contain any of the points in $\{a_1, \ldots, a_k\}$.

Note that for each j, at most 4 colors are used on intervals in $A_j \cup B_j$. It follows that for each $c \in \{a_1, \ldots, a_k\} \cup \{b_1, \ldots, b_k\}$, at most 4 colors are used on intervals in X^c. It remains to check that the coloring is proper. Note that the colors used on A_j are disjoint from the colors used on B_j. Since the coloring is proper on A_j and on B_j, it follows that the coloring is proper on $A_j \cup B_j$. Moreover, if $x \in A_i \cup B_i$ and $y \in A_j \cup B_j$ overlap, it follows that $|i - j| \leq 1$ and y overlap, since each interval in X meets exactly one of the sets in $\{S_1, \ldots, S_k\}$.

Suppose that $x \in A_i \cup B_i$ and $y \in A_j \cup B_j$ overlap. If $i = j$, then x and y receive different colors since the coloring is proper on $A_i \cup B_i$. Hence, we may assume that $j = i + 1$. Note that $a_j \in y$, since otherwise a_j would separate x and y. It follows that $y \in A_j$. If y has height 0 in A_j, then the color assigned to y is not used for intervals in $A_i \cup B_i$, and hence x and y receive different colors. If y has height 1 in A_j, then the color β assigned to y is used only for the intervals in $A_i \cup B_i$ that have depth 1 in B_i. Suppose for a contradiction that x also receives color β. Since x has depth 1 in B_i, there exists $x' \in B_i$ with $x < x'$ in the endpoint order. Similarly, since y has height 1 in A_j, there exists $y' \in A_j$ with $y' < y$ in the endpoint order. Moreover, $l(x') < b_i < l(y)$ and $r(x') < a_3 < r(y')$, and, therefore, $x' < y'$ in the endpoint order. But then $\{x, x', y', y\}$ is a 4-clique in $G(X)$ since $x < x' < y' < y$ in the endpoint order and x and y overlap. □

Example 4.2. The complement of the cycle on 7 vertices, denoted $\overline{C_7}$, is the overlap graph of a family of intervals that satisfies the hypotheses of Lemma 4.1 with carefully chosen points a_1 and b_1 (see Fig. 2). Consequently, Lemma 4.1 cannot be improved by more than one color.

Our next task is to explore the structure of segments. A *segment* of a family X is an inclusion-maximal interval in the set of all centers of 2-cliques in X.

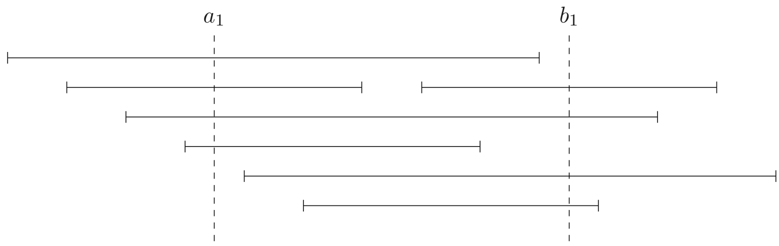


Fig. 2 $\overline{C_7}$ as the overlap graph of a family of intervals

Lemma 4.3. *Let X be a family of intervals. If $[a,b]$ and $[c,d]$ are overlapping segments of X with $a < c < b < d$, then there exists $x \in X$ with $l(x) \in [a,c)$ and $r(x) \in (b,d]$.*

Proof. Let y_1 and y_2 be overlapping intervals in X with $y_1 < y_2$ and center $[a,b]$. Let z_1 and z_2 be overlapping intervals in X with $z_1 < z_2$ and center $[c,d]$. Note that $l(y_2) = a$ and $r(z_1) = d$. We claim that either $r(y_2) \in (b,d]$ or $l(z_1) \in [a,c)$. Because $r(y_2) > b$ and $l(z_1) < c$, failure requires $r(y_2) > d$ and $l(z_1) < a$. But then we have $l(z_1) < l(y_2) = a < d = r(z_1) < r(y_2)$, which implies that z_1 and y_2 are overlapping intervals in X with center $[a,d]$, contradicting that $[a,b]$ and $[c,d]$ are segments. Hence, either y_2 or z_1 is as required. □

Lemma 4.4. *Let X be a family of intervals. If $u_1, \ldots, u_t$ are overlapping segments of X with $u_1 < u_2 < \cdots < u_t$, then $[l(u_t), r(u_1)]$ is the center of a $(t+1)$-clique in X.*

Proof. For $1 \le j < t$, apply Lemma 4.3 to the segments u_j and u_{j+1} to obtain $z_j \in X$ with $l(z_j) \in [l(u_j), l(u_{j+1}))$ and $r(z_j) \in (r(u_j), r(u_{j+1})]$. Of the overlapping pair of intervals in X whose center is u_1, let z_0 be the leftmost in the endpoint order. Similarly, of the overlapping pair of intervals in X whose center is u_t, let z_t be the rightmost in the endpoint order. It follows that $l(z_0) < l(z_1) < \cdots < l(z_t) < r(z_0) < r(z_1) < \cdots < r(z_t)$ and so $\{z_0, \ldots, z_t\}$ is a $(t+1)$-clique in X with center $[l(u_t), r(u_1)]$. □

As a consequence of Lemma 4.4, if X has clique number k and U is the family of segments of X, then U has clique number at most $k-1$. Moreover, by definition, each interval in U is inclusion-maximal. Hence, the endpoint order on U is a chain. If X has clique number at most 3, then every component of the overlap graph of U is a chain.

We need the following lemma due to Gyárfás [9].

Lemma 4.5. *Let X be a of intervals such that $G(X)$ is connected, let x_0 be the interval in X that minimizes $l(x_0)$, and for each $k \ge 0$, let X_k be the set of all intervals at distance k from x in $G(X)$. Let k be a positive integer, and let $[a,b]$ be an interval such that $[a,b] \subseteq \bigcup_{x \in X_k} x$. If $z \in S_{k-1}$ and one endpoint of z is in $[a,b]$, then the other endpoint of z is outside $[a,b]$.*

Our next lemma ties together the two separate coloring strategies given by Theorem 3.2 and Lemma 4.1 and is at the heart of our proof. The *clean part* of a family of intervals X is the set of intervals in X that are not contained in a segment of X. Note that the clean part of a family of intervals is clean. In the following, we fix disjoint color sets A and B of sizes 10 and 9, respectively.

Lemma 4.6. *Let X be a family of intervals such that no interval is contained in the center of a 3-clique, $G(X)$ is connected, and $\omega(G(X)) \leq 3$. Let x_0 be the interval in X that minimizes $l(x_0)$, and for $k \geq 0$, let X_k be the set of intervals at distance k from x_0. Let Y_k be the clean part of X_k, and let Z_k be the complement $X_k - Y_k$.*

For each nonnegative integer k, there are a set P_k of points and a proper $(A \cup B)$-coloring of $G(X_0 \cup \ldots \cup X_k)$ with the following properties.

(1) If $j > k$ and $x \in X_j$, then x and P_k are disjoint.
(2) Every interval in Z_k contains a point in P_k.
(3) The colors used on Y_k are contained in a subset A' of A with $|A'| \leq 5$.
(4) The colors used on Z_k are contained in B.
(5) Let I be an inclusion-maximal interval in $\mathbb{R} - P_k$. There exists a subset B' of B with $|B'| \leq 5$ such that every interval in Z_k that overlaps I has a color in B', and there is a color $\beta \in B'$ such that z overlaps I on the left whenever $z \in Z_k$ overlaps I and has color β.

Proof. For $k = 0$, we let $P_0 = \emptyset$ and color the single interval x_0 in X_0 with an arbitrary color in A. Since $Y_0 = \{x_0\}$ and $Z_0 = \emptyset$, conditions (1)–(5) are satisfied.

For $k \geq 1$, we obtain a set of points P_{k-1} and a proper $(A \cup B)$-coloring of $G(X_0 \cup \ldots \cup X_{k-1})$ with conditions (1)–(5) by induction. We first extend the coloring to $G(X_0 \cup \ldots \cup X_k)$. Note that an interval in X_k overlaps only with intervals in $X_{k-1} \cup X_k$. Since property (3) implies that at most 5 colors are used on intervals in Y_{k-1} and $|A| = 10$, there is a set A' of 5 colors in A, none of which appears on intervals in Y_{k-1}. Since Y_k is clean, Theorem 3.2 implies that Y_k has a proper 5-coloring. We use the colors in A' to color Y_k.

Every interval in Z_k is contained in a segment of X_k. Let $u_1, \ldots, u_s$ be the segments of X_k, indexed so that $u_1 < \cdots < u_s$ in the endpoint order. For each segment u_j, we define a *left-pin* a_j and a *right-pin* b_j. Let ε be a positive real number that is less than the minimum distance between two endpoints of intervals in X. When there are intervals in X_{k-1} that overlap u_j on the left, we define a_j to be $\max\{r(x)\colon\ x \in X_{k-1}$ and x overlaps u_j on the left$\}$. When there are no such intervals, we define a_j to be $r(u_{j-1}) + \varepsilon$ when u_{j-1} exists and overlaps u_j and $l(u_j) + \varepsilon$ otherwise. Similarly, when there are intervals in X_{k-1} that overlap u_j on the right, we define b_j to be $\min\{l(x)\colon\ x \in X_{k-1}$ and x overlaps u_j on the right$\}$. When there are no such intervals, we define b_j to be $l(u_{j+1}) - \varepsilon$ when u_{j+1} exists and overlaps u_j and $r(u_j) - \varepsilon$ otherwise.

Note that no pin is in more than one segment. If some pin $c \in \{a_1, \ldots, a_s, b_1, \ldots, b_s\}$ were contained in u_j and u_{j+1}, it follows from the definition of c that there is an interval $x \in X_{k-1}$ with c as an endpoint. Moreover, Lemma 4.4 implies that there is a 3-clique $\{x_1, x_2, x_3\}$ in X_k whose center is the same as the center of $\{u_j, u_{j+1}\}$.

Since c is in the center of $\{x_1,x_2,x_3\}$, Lemma 4.5 implies that the other endpoint of x is outside $x_1 \cup x_2 \cup x_3$, and so x overlaps each interval in $\{x_1,x_2,x_3\}$. Therefore, $\{x_1,x_2,x_3,x\}$ is a 4-clique in X, a contradiction.

Next, we argue that every interval $z \in Z_k$ contains some pin in its interior. Since $z \in Z_k$, it follows that z is contained in some segment u_j. Since z is at distance k from x_0 in $G(X)$, it follows that z overlaps with an interval x at distance $k-1$ from x_0 in $G(X)$. Suppose that x overlaps z to the left, so that $r(x)$ is in the interior of z. We claim that z contains the left-pin of u_j. Since $x \in X_{k-1}$, it follows that $a_j \geq r(x) > l(z)$. Let x' be the interval in X_{k-1} whose right endpoint is a_j, and let x_1 and x_2 be the intervals in X_k whose center is the segment u_j. Lemma 4.5 implies that the left endpoint of x' is outside $x_1 \cup x_2$, which implies that $\{x',x_1,x_2\}$ is a 3-clique in X. Since z is contained in the center of x_1 and x_2 and z is not contained in a 3-clique of X, it must be that z is not contained in x', which implies that $a_j = r(x') < r(z)$. Hence, $l(z) < a_j < r(z)$. Similarly, if x overlaps z on the right, then z contains the right-pin of u_j.

For each j with $1 \leq j \leq s$, let $S_j = \{a_j, b_j\}$. Since an interval $z \in Z_k$ is contained in some segment u_j and u_j contains only the pins a_j and b_j, it follows that z has nonempty intersection with exactly one of the sets in $\{S_1,\ldots,S_s\}$. We claim that if c is a pin, then $\omega(G(Z_k^c)) \leq 2$. This is immediate if c is an endpoint of an interval in X_{k-1}. Otherwise, let c' be the other pin associated with the segment containing c, and note that $Z_k^c \subseteq Z_k^{c'}$. If c' is an endpoint of an interval in X_{k-1}, then we have $\omega(G(Z_k^c)) \leq \omega(G(Z_k^{c'})) \leq 2$. If neither c nor c' is the endpoint of an interval in X_{k-1}, then it must be that $Z_k^c = Z_k^{c'} = \emptyset$. Therefore, every subset of Z_k satisfies the hypotheses of Lemma 4.1 with respect to the points $a_1,\ldots,a_s$ and $b_1,\ldots,b_s$.

It remains to color Z_k. Let I be an inclusion-maximal interval of $\mathbb{R} - P_{k-1}$, and let L be the set of intervals in Z_k that are contained in I. Since intervals in distinct inclusion-maximal intervals of $\mathbb{R} - P_{k-1}$ do not overlap, we may color L independently of the rest of Z_k.

By property (5), there exists $B_0 \subseteq B$ with $|B_0| \leq 5$ such that every interval in Z_{k-1} that overlaps I has a color in B_0, and there is a distinguished color $\alpha \in B_0$ such that if $z \in Z_{k-1}$ overlaps I and has color α, then z overlaps I on the left. Let $B' = B - B_0 \cup \{\alpha\}$, and note that $|B'| = 5$.

Using the colors in B' and the distinguished color α, apply Lemma 4.1 to color L. We claim that the coloring remains proper. If not, then there are intervals $z \in L$ and $z' \in X_{k-1}$ that overlap and have the same color. Since the color of z is in B', the color of z' is also in B', which implies that $z' \in Z_{k-1}$. By property (2), we have that z' contains a point in P_{k-1}, and it follows that z' overlaps I. Hence, property (5) implies that the color of z' is in B_0. Since $B_0 \cap B' = \{\alpha\}$, it follows that the common color of z and z' is α. It now follows that z' overlaps I on the left. Since z' is an interval in X_{k-1} that overlaps z on the left, it follows that z contains the left-pin a_j of the segment u_j containing z, contradicting that each interval in L with color α is disjoint from $\{a_1,\ldots,a_s\}$.

We have obtained a proper $(A \cup B)$-coloring of $G(X_0 \cup \cdots \cup X_k)$. Let P_k be the union of P_{k-1} and the points in $\{a_1, \ldots, a_s, b_1, \ldots, b_s\}$ that are endpoints of intervals in X_{k-1}. It remains to check that the coloring and P_k satisfy properties (1)–(5). Note that if x has distance $k-1$ from x_0 and x' has distance at least $k+1$, then x' does not contain either endpoint of x. Since every point in $P_k - P_{k-1}$ is an endpoint of some interval in X_{k-1}, it follows that no interval in X_j for $j > k$ contains a point in P_k, which implies property (1). If $z \in Z_k$, then z overlaps some interval $x \in X_{k-1}$. Let u be the segment of X_k containing z. If x overlaps z on the left, then z contains the left-pin of u, which is the endpoint of an interval in X_{k-1}. Otherwise, if x overlaps z on the right, then z contains the right-pin of u, which is the endpoint of an interval in X_{k-1}. In either case, z contains a point in P_k, and therefore property (2) is satisfied. It is clear from our coloring that properties (3) and (4) are satisfied.

Let I be an inclusion-maximal interval in $\mathbb{R} - P_k$. Since $P_{k-1} \subseteq P_k$, it follows that I is contained in an inclusion-maximal interval I' in $\mathbb{R} - P_{k-1}$. Let L be the set of intervals in Z_k that are contained in I', and let B' be the set of 5 colors in B that are used to color intervals in L. Clearly, every interval in Z_k that overlaps I has a color in B'. Let c be the right endpoint of I. By Lemma 4.1, at most 4 colors are used on intervals in L^c. It follows that there is a color $\beta \in B'$ such that every interval in L with color β that overlaps I does so on the left. It follows that property (5) is satisfied. □

With Lemma 4.6, we are now able to complete our upper bound on the chromatic number of a circle graph with clique number at most 3.

Proof of Theorem 1.2. We may assume that $G(X)$ is connected. Let x_0 be the interval in X that minimizes $l(x_0)$, and for $k \geq 0$, let X_k be the set of intervals that are at distance k from x_0 in $G(X)$.

Note that no interval in X_k is contained in the center of a 3-clique of X_k. This is immediate if $k = 0$ since $X_0 = \{x_0\}$. For $k \geq 1$, if some interval x were contained in the center of a 3-clique $\{x_1, x_2, x_3\}$ in X_k, then there is an interval x' in X_{k-1} that overlaps x, and Lemma 4.5 would imply that $\{x_1, x_2, x_3, x'\}$ is a 4-clique in $G(X)$, a contradiction.

Therefore, Lemma 4.6 implies that $\chi(G(X_k)) \leq 19$. Using disjoint color sets for $X_0 \cup X_2 \cup \cdots$ and $X_1 \cup X_3 \cup \cdots$, we have that $\chi(G(X)) \leq 38$. □

Addendum. Subsequent to our current work but prior to publication, we learned that Gleb Nenashev has improved on Theorem 1.2 by showing that $\chi(G) \leq 30$ when G is a K_4-free circle graph.

Acknowledgements The first author's research is supported in part by NSF Grant DMS-0965587 and by Grant 09-01-00244-a of the Russian Foundation for Basic Research.

The second author acknowledges support of the National Science Foundation through a fellowship funded by Grant EMSW21-MCTP, "Research Experience for Graduate Students" (NSF DMS 08-38434).

References

1. A.A. Ageev, A triangle-free circle graph with chromatic number 5. Discrete Math. **152**(1–3), 295–298 (1996)
2. A. Bouchet, Un algorithme polynomial pour reconnatre les graphes d'alternance. (french) [a polynomial algorithm for recognizing circle graphs]. Can. J. Math. **11**(16), 34–38 (1959)
3. H. de Fraysseix, A characterization of circle graphs. Eur. J. Comb. **5**, 223–238 (1983)
4. P. Erdős, Graph theory and probability. C. R. Acad. Sci. Paris Sr. I Math. **300**, 569–572 (1985)
5. S. Even, A. Itai, Queues, stacks and graphs. *Theory of Machines and Computations* (Academic Press, New York, 1971), pp. 71–86
6. M. Garey, D. Johnson, G. Miller, C. Papadimitriou, The complexity of coloring circular arcs and chords. SIAM J. Alg. Disc. Method **1**, 216–227 (1980)
7. F. Gavril, Algorithms for a maximum clique and a maximum independent set of a circle graph. Networks **3**, 261–273 (1973)
8. M.C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs* (Academic Press, New York, 1980)
9. A. Gyárfás, On the chromatic number of multiple interval graphs and overlap graphs. Discrete Math. **55**(2), 161–166 (1985)
10. A. Gyárfás, Corrigendum: "On the chromatic number of multiple interval graphs and overlap graphs." Discrete Math. **62**(3), 333 (1986)
11. A. Gyárfás, Problems from the world surrounding perfect graphs. Zastos. Mat. **19**, 413–441 (1988)
12. A. Gyárfás, J. Lehel, Covering and coloring problems for relatives of intervals. Discrete Math. **55**, 167–180 (1985)
13. I.A. Karapetyan, Chordal graphs. Mat. Voprosy Kibernet. Vychisl. Tekhn. (14), 6–10 (1985)
14. J.M. Keil, L. Stewart, Approximating the minimum clique cover and other hard problems in subtree filament graphs. Discrete Appl. Math. **154**, 1983–1995 (2006)
15. A. Kostochka, Coloring intersection graphs of geometric figures with a given clique number. In *Towards a Theory of Geometric Graphs*, vol. 342 of Contemp. Math. (American Mathematical Society, Providence, RI, 2004), pp. 127–138
16. A. Kostochka, J. Kratochvíl, Covering and coloring polygon-circle graphs. Discrete Math. **163**(1–3), 299–305 (1997)
17. A.V. Kostochka, On upper bounds for the chromatic numbers of graphs. Trudy Instituta Mathematiki **10**, 204–226 (1988) [in Russian]
18. J. Mycielski, Sur le coloriage des graphs. Colloq. Math. **3**, 161–162 (1955)

Counting Large Distances in Convex Polygons: A Computational Approach

Filip Morić and David Pritchard

Abstract In a convex n-gon, let $d_1 > d_2 > \cdots$ denote the set of all distances between pairs of vertices, and let m_i be the number of pairs of vertices at distance d_i from one another. Erdős, Lovász, and Vesztergombi conjectured that $\sum_{i \le k} m_i \le kn$. Using a new computational approach, we prove their conjecture when $k \le 4$ and n is large; we also make some progress for arbitrary k by proving that $\sum_{i \le k} m_i \le (2k-1)n$. Our main approach revolves around a few known facts about distances, together with a computer program that searches all distance configurations of two disjoint convex hull intervals up to some finite size. We thereby obtain other new bounds, such as $m_3 \le 3n/2$ for large n.

1 Introduction

Given a set S of n points in the plane, let $d_1 > d_2 > \cdots$ be the set of all distances between pairs of points in S. It was shown by Hopf and Pannwitz in 1934 [5] that the distance d_1 (the diameter of S) can occur at most n times, which is tight (e.g., for a regular polygon of odd order). In 1987, Vesztergombi [6] showed that the second-largest distance, d_2, can occur at most $\frac{3}{2}n$ times; she subsequently [7] considered the version of the problem when the points are in convex position and showed that in this case the number of second-largest distances is at most $\frac{4}{3}n$. She also showed that both results are tight up to additive constants.

F. Morić
Ecole Polytechnique Fédérale de Lausanne, Chair of Combinatorial Geometry,
EPFL SB IMB DCG, MA C1 585 (Bâtiment MA), CH-1015 Lausanne, Switzerland

D. Pritchard (✉)
Centre for Education in Math and Computing, University of Waterloo,
200 University Ave W, N2L 3G1 Waterloo, ON, Canada
e-mail: daveagp@gmail.com

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_22,

Let m_i denote the number of times that d_i occurs. It is known that $m_k \leq 2kn$ [6], and moreover that $m_k \leq kn$ for point sets in convex position [7], while the following open conjecture would imply $m_k \leq 2n$.

Conjecture 1.1 (Erdős, Moser [2, 7]). The number of unit distances generated by n points in convex position cannot exceed $2n$.

A lower bound of $2n-7$ for this conjecture is known due to Edelsbrunner and Hajnal [3].

For the rest of the chapter, we consider only point sets in convex position. One natural question is to find how large $m_{\leq k} := \sum_{i \leq k} m_i$, i.e., the number of *top-k* distances, can be in terms of n. The conjectured value is

Conjecture 1.2 (Erdős, Lovász, Vesztergombi [4]). The number of top-k distances generated by n points in convex position is at most kn; i.e., $m_{\leq k} \leq kn$.

Odd regular polygons prove $m_{\leq k} = kn$ is possible. In [4], the bound $m_{\leq k} \leq 3kn$ is proven, and $m_{\leq 2} \leq 2n$ was shown in [7], verifying Conjecture 1.2 for $k = 2$.

In this chapter, we give improved upper bounds on m_k and $m_{\leq k}$ for convex point sets, and more generally bounds for sums of the form $\sum_{t \in T} m_t$. Our first result is the following.

Theorem 3. *For any $k \geq 1$, the number of top-k distances generated by n points in convex position is at most $(2k-1)n$; i.e., $m_{\leq k} \leq (2k-1)n$.*

Thus, we close about half of the gap toward Conjecture 1.2.

Next, by combining several known conditions on distances for convex point sets, and by using a computer program to carry out an exhaustive search on a finite abstract version of the problem, we prove the following.

Theorem 4. *The distances generated by n points in convex position satisfy the following bounds, for large enough n:*

- $m_{\leq 3} \leq 3n, m_{\leq 4} \leq 4n$;
- $m_3 \leq \frac{3}{2}n, m_4 \leq \frac{13}{8}n$;
- $m_1 + m_3 \leq 2n, m_2 + m_3 \leq \frac{9}{4}n$.

In particular, we verify Conjecture 1.2 for $k \leq 4$ and n large. For m_3 and $m_2 + m_3$, the bound is as good as can be obtained by our abstract version of the problem, as witnessed by periodic patterns achieving $m_3 = \frac{3}{2}n$ and $m_2 + m_3 = \frac{9}{4}n$, but we do not know if any convex polygon can realize these distances; we elaborate in Sect. 6.

The proof of Theorem 4 uses a computer program to make certain types of automatic deductions, as well as the following lemma to eliminate long distances "near" the boundary.

Lemma 1.5. *For any $k \geq 1$ and $\ell \geq 0$, there is a constant $C(k,\ell)$ such that the following holds: In a convex polygon, if there are ℓ or fewer vertices between some vertices a and b such that $|ab| \geq d_k$, then the number of top-k distances satisfies $m_{\leq k} \leq n + C(k,\ell)$.*

The detailed bound we obtain is of the form $C(k,\ell) = O(k^2(k+\ell)^2)$. In an earlier version of this chapter,[1] we proved results like "$m_{\leq 3} \leq 3n + O(1)$," which are weaker for large n but better for small n, using the following alternative lemma.

Lemma 1.6. *For any $k \geq 1$ and $\ell \geq 0$, there is a constant $C'(k,\ell)$ such that the following holds. In a convex polygon, at most $C'(k,\ell)$ diagonals ab have both (i) ℓ or fewer vertices between a and b and (ii) $|ab| \geq d_k$.*

In the latter, $C'(k,\ell) = O(k\ell^2)$. We do not think either lemma is tight.

In Sect. 2, we describe *levels*, a key element in our approach. In Sect. 3, we collect geometric facts used by the algorithm. We prove Lemma 1.5 in Sect. 3.1. The proof of our main result, Theorem 4, consists of the algorithmic approach described in Sect. 4 together with our computational results stated in Sect. 5. We conclude with suggestions for future work.

2 Levels

We use the term *diagonal* to mean any line segment connecting two points of S, including sides of the convex hull of S. We will partition the diagonals into n *levels* in the following way. Let $S = \{a_1, a_2, \ldots, a_n\}$ be the vertex set of our convex polygon, ordered clockwise. Then *level i* is the set of diagonals

$$L_i := \{a_j a_k \mid j + k \equiv i \bmod n\},$$

where the index i can be taken modulo n. Equivalently, consider an auxiliary regular n-gon $b_1 b_2 \ldots b_n$; then two diagonals $a_i a_j$ and $a_k a_l$ lie in the same level when the corresponding segments $b_i b_j$ and $b_k b_l$ are parallel. We illustrate this in Fig. 1a.

Levels are used in the following way to prove Theorem 3 (i.e., $m_{\leq k} \leq (2k-1)n$).

Proof of Theorem 1.3. In the next section, we prove Lemma 3.5: In any level, there are at most $2k-1$ diagonals of length $\geq d_k$. Since there are at most n levels, we are done. □

3 Geometric Facts

To begin this section, we collect four geometric facts from the literature [1, 4, 7], which will be used in our computer program. For completeness, we include the proofs. The first two facts were used in [4, 7].

Fact 3.1. If $abcd$ is a convex quadrangle, then $|ab| + |cd| < |ac| + |bd|$.

[1] http://arxiv.org/abs/1103.0412v1.

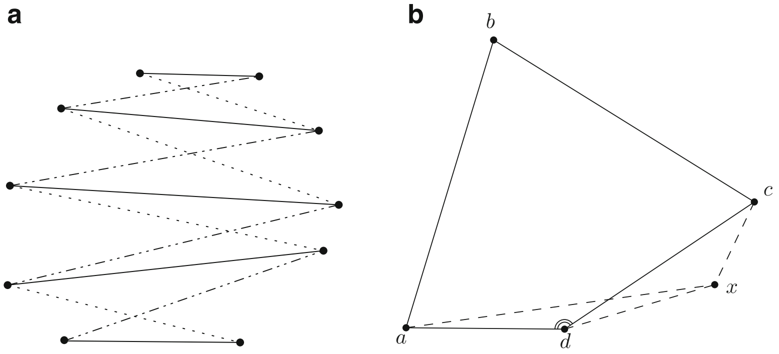


Fig. 1 (**a**) Three consecutive levels of diagonals in a convex decagon. (**b**) Proof of Fact 3.2

Proof. Let p be the intersection point of the diagonals ac, bd. Then, by the triangle inequality,

$$|ab| + |cd| < |ap| + |bp| + |cp| + |dp| = |ac| + |bd|. \qquad \square$$

Fact 3.2. If a,b,c,d are vertices of a convex polygon in clockwise order, then at least one of these four cases must occur:

- $|ax| > |ad|$ for all vertices x of the polygon between c and d, including c;
- $|bx| > |bc|$ for all vertices x of the polygon between c and d, including d;
- $|cx| > |bc|$ for all vertices x of the polygon between a and b, including a;
- $|dx| > |ad|$ for all vertices x of the polygon between a and b, including b.

Proof. Since the sum of the angles of quadrilateral $abcd$ is 2π, at least one angle is nonacute. Without loss of generality, let $\angle cda \geq \frac{\pi}{2}$. Then for any vertex x of the polygon between c and d, we have that $\angle xda \geq \angle cda \geq \frac{\pi}{2}$, and, thus, $|ax| > |ad|$ (see Fig. 1b). $\square$

The special case $i = j$ of the following fact appears in [4].

Fact 3.3. If a,b,c,d are vertices of a convex polygon listed in clockwise order, such that $|bc| \geq d_i$ and $|ad| \geq d_j$, where d_i and d_j are the ith- and jth-largest distances among vertices of the polygon, then either between a and b or between c and d there are no more than $i + j - 3$ other vertices of the polygon.

Proof. Let's denote without loss of generality $a = a_1, b = a_x, c = a_y, d = a_z$. We will show $\min\{x-1, z-y\} \leq i + j - 2$, which proves the lemma. We use induction on $i + j$ see Fig. 2. The base case $i = j = 1$ amounts to saying that any two noncrossing d_1's must share a vertex, which follows by Fact 3.1.

For the inductive step, we apply Fact 3.2. Suppose that the first of the four cases happens, so $d' := a_{z-1}$ satisfies $|ad'| > |ad|$; the other cases are similar. Consequently, $|ad'| \geq d_{j-1}$. By induction, $\min\{x-1, (z-1)-y\} \leq i + (j-1) - 3$, from which the desired result follows (Fig. 2). $\square$

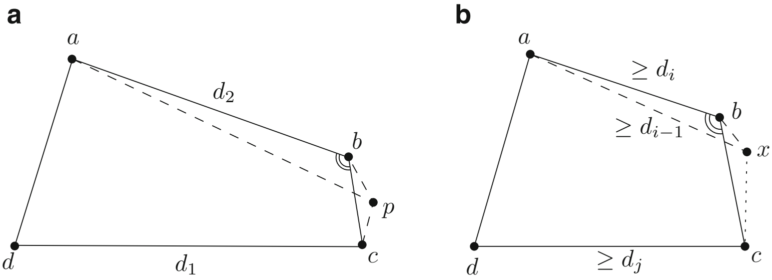


Fig. 2 (**a**) Proof of Fact 3.3, base case $i = 2$, $j = 1$; (**b**) proof of Fact 3.3, inductive step

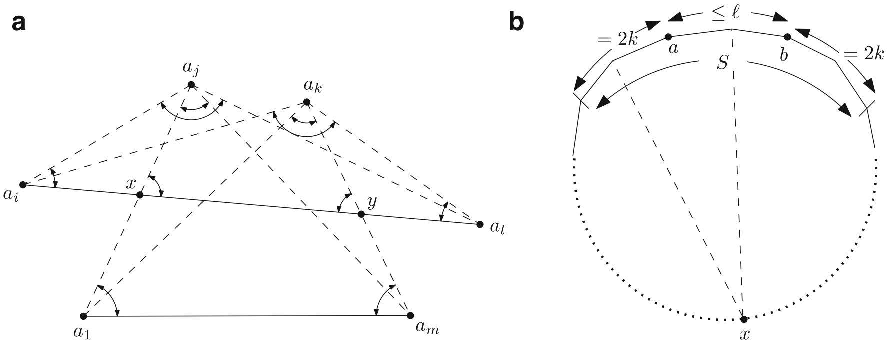


Fig. 3 (**a**) Proof of Fact 3.4; (**b**) proof of Lemma 1.5

The following is a strengthening of a result of Altman, obtained by removing all nonessential conditions from the hypothesis of [1, Lemma 1] but using the same proof. (He considered only the case where $|a_1a_m| = d_1$.)

Fact 3.4. Let $a_1 \dots a_n$ be a convex polygon. If $1 \leq i < j \leq k < \ell < m$ and $|a_1a_m| \geq \max\{|a_1a_k|, |a_ja_m|\}$, then $|a_ia_\ell| > \min\{|a_ia_k|, |a_ja_\ell|\}$.

Proof. Suppose for the sake of contradiction that $|a_ia_\ell| \leq \min\{|a_ia_k|, |a_ja_\ell|\}$. Denote by x and y the points where a_1a_j and a_ma_k intersect a_ia_ℓ (see Fig. 3a). Repeatedly using the fact that when s, s' are two sides of a triangle, $|s| > |s'|$ iff the angle opposite s is larger than the angle opposite s', we have

$$\angle a_jxa_\ell + \angle a_kya_i > \angle a_ja_ia_\ell + \angle a_ka_\ell a_i \geq \angle a_ia_ja_\ell + \angle a_\ell a_ka_i$$
$$> \angle a_1a_ja_m + \angle a_1a_ka_m \geq \angle a_ja_1a_m + \angle a_ka_ma_1 \,.$$

However, $\angle a_jxa_\ell + \angle a_kya_i = \angle a_ja_1a_m + \angle a_ka_ma_1$, which gives a contradiction. □

3.1 Counting Lemmas

First, we complete the proof of Theorem 3, using Fact 3.3.

Lemma 3.5. *In any level, there are at most* $2k-1$ *diagonals of length* $\geq d_k$.

Proof. Without loss of generality (by relabeling), we consider the level L_0. The diagonals of this level are $a_j a_{-j}$, with indices modulo n, for $0 < j < n/2$. Let $m > 0$ (resp., M) be the minimal (resp., maximal) j such that $|a_j a_{-j}| \geq d_k$. Then, by Fact 3.3, we see that $M - m - 1 \leq k + k - 3$. So the number of top-k diagonals in L_0 is bounded by $|\{m, m+1, \ldots, M\}| = M - m + 1 \leq 2k - 1$, which gives the corollary. □

Next, we give the proof of Lemma 1.5, which is needed in order to argue that our computational approach is correct.

Proof. We want to show that if $|ab| \geq d_k$, and a and b are separated by at most ℓ vertices, then the number of top-k distances satisfies $m_{\leq k} \leq n + O(k^2(k+\ell)^2)$. Let S be the interval obtained from this $[a,b]$ by extending onto $2k$ further points in both directions. By Fact 3.3, all edges of length $\geq d_k$ have at least one endpoint in S. Note $|S| = O(k+\ell)$.

We will show an upper bound of $n + O(k^2(k+\ell)^2)$ on the number of edges sx of length $\geq d_k$, with $s \in S, x \in V \backslash S$. This will complete the proof since the only other top-k distance edges must lie with both endpoints in S, and there are at most $O(k+\ell)^2$ such edges.

The key observation is that in the bipartite graph between S and $V \backslash S$ consisting of these edges, all but a constant number of vertices in $V \backslash S$ have degree 1. Specifically, if $sx, s'x$ are both edges in this graph, then the location of x is uniquely determined by $s, s', |sx|$, and $|s'x|$; it follows that $\sum_x \binom{\deg(x)}{2}$ is at most $O((k+\ell)^2 k^2)$, and consequently $\sum_{x:\deg(x)>1} \deg(x) = O((k+\ell)^2 k^2)$. We are then done by counting the endpoints of degree-1 vertices, of which there are at most n. □

4 The Algorithm

The algorithm we use to prove Theorem 4 examines distances among finite configurations of points in the plane. Informally, we examine all possible configurations of a bounded size, where a configuration includes all occurrences of top-k distances in a few consecutive levels, and we try to establish that not too many top-k distances can occur per level, averaged over a small interval of levels. Thus, ultimately, the argument in our proof decomposes any global point set into local configurations of bounded size.

4.1 The Goal

Our computational goal will be to bound the number of long distances that can occur in a consecutive sequence of several levels. We begin by reproving (for large n) Vesztergombi's result on counting the second-largest distances; it illustrates the type of computational result we need.

Proposition 4.1. *We have $m_2 \leq \frac{4}{3}n$ for large enough n.*

Proof. We prove the theorem for $n \geq 3 \cdot C(16,2)$ with C as in Lemma 1.5. Let a *special diagonal* be a diagonal of length d_2 or longer, whose endpoints are separated by at most 16 vertices. If there is any special diagonal, we are done by Lemma 1.5. So we may assume there are no special diagonals.

Using our computer program, we establish the following lemma.

Lemma 4.2. *In every point set S without special diagonals, for every level i, at least one of the following is true:*

- *at most* $1 = \lfloor 1 \cdot \frac{4}{3} \rfloor$ *diagonal in level i has length* d_2*;*
- *at most* $2 = \lfloor 2 \cdot \frac{4}{3} \rfloor$ *diagonals in levels i and* $i+1$ *have length* d_2*;*
- *at most* $4 = \lfloor 3 \cdot \frac{4}{3} \rfloor$ *diagonals in levels* $i,\ldots,i+2$ *have length* d_2*;*
- *at most* $5 = \lfloor 4 \cdot \frac{4}{3} \rfloor$ *diagonals in levels* $i,\ldots,i+3$ *have length* d_2*.*

Now let's see how this gives the desired result. Taking $i = 1$, the four cases above establish that for some $1 \leq \gamma_1 \leq 4$, the number of d_2's in levels $1,\ldots,\gamma_1$ is at most $\frac{4}{3}\gamma_1$. Applying the same logic to $i = \gamma_1 + 1$, we get that there is some $1 \leq \gamma_2 \leq 4$ such that the number of d_2's in levels $\gamma_1+1,\ldots,\gamma_1+\gamma_2$ is at most $\frac{4}{3}\gamma_2$.

We continue to further define γ_i's in the same way until $\sum_{i=1}^{x}\gamma_i \equiv \sum_{i=1}^{y}\gamma_i \pmod{n}$ for some $x < y$. Summing a contiguous subset of these bounds, the number of d_2's in levels from $1+\sum_{i=1}^{x}\gamma_i$ to $\sum_{i=1}^{y}\gamma_i$ is at most $\frac{4}{3}$ per level on average. But this sum counts each of the n levels an equal number of times, so the number of d_2's overall is at most $\frac{4}{3}n$. □

The computer program's goal is thus to prove a general version of Lemma 4.2: Given a *target ratio* α and *target distances* (a subset of $\{d_1,d_2,\ldots,d_k\}$), find a constant m so that every level i admits $1 \leq m' \leq m$ such that $\leq m' \cdot \alpha$ target lengths occur in levels $i,\ldots,i+m'$. The program searches for a point set with $> \alpha$ target diagonals in level 1, $> 2\alpha$ in level 2, etc. If the search terminates, the above proof shows the number of target distances is $\leq \alpha n$. The hypothesis that no special diagonals exist is used only indirectly by the program, explained below.

Our algorithm works with *configurations* consisting of two disjoint intervals of points, and an assignment of a distance from $\{d_1,d_2,\ldots,d_k,\text{“}<d_k\}$ to each diagonal spanning the two intervals. We thereby obtain analogues of Lemma 4.2 by checking all possible configurations up to some finite size. For this to work, Fact 3.2 is crucial since it implies that all of the top-k distances in ℓ consecutive levels have all of their endpoints in two intervals of bounded size. We use an incremental branch-and-bound search: It exhaustively searches all possibilities, but in an efficient way

where large sections of the search space can be eliminated at once. Each individual step of the algorithm corresponds to an application of one of the Facts 3.1–3.4.The lack of special diagonals allows us to focus on *disjoint* interval pairs. The Java implementation is available at

http://sourceforge.net/projects/convexdistances/.

4.2 Configurations

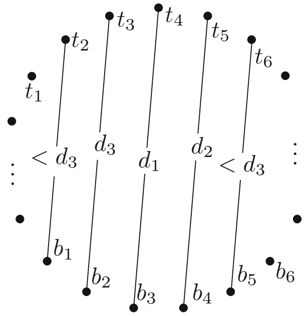
In more detail, our algorithm maintains a set of *configurations*. Each configuration has two disjoint intervals of points from S; then for each diagonal generated by one point from each interval, the configuration stores a set of possible values for the distance between those two points. Arbitrarily name one interval the *top* and denote its points as $\{t_i\}_i$, with t_{i+1} following t_i in clockwise order, and name the other interval the *bottom* with points $\{b_i\}_i$, and b_{i-1} following b_i in clockwise order. Then we denote the set of possible distances between t_i and b_j as $D[i,j]$; in each configuration $D[i,j]$ is a subset of $\{1,2,\ldots,k,\infty\}$, where $x \in D[i,j]$ means that d_x is a possible value for the distance $|t_ib_j|$, while $\infty \in D[i,j]$ means that it is possible for $|t_ib_j|$ to be shorter than d_k. (So typical steps in our program use special cases to reason with "d_∞" distances correctly.) Reiterating, a configuration consists of a top interval of indices, a bottom interval of indices, and for each top-bottom pair a subset of $\{1,2,\ldots,k,\infty\}$.

We assume that t_ib_j is in level number $j-i$ (modulo n), which is without loss of generality. To gain some intuition and exhibit the notation, it is helpful to look at a couple of examples. Our examples will be drawn from actual point sets and therefore each $D[i,j]$ will be just a singleton, in contrast to the larger sets $D[i,j]$ typically occurring in the algorithm. The first example, shown in Fig. 4, is a regular polygon of odd order. The second example, shown in Fig. 5, exhibits the extremal construction of Vesztergombi for second distances [7].

4.3 Methodology

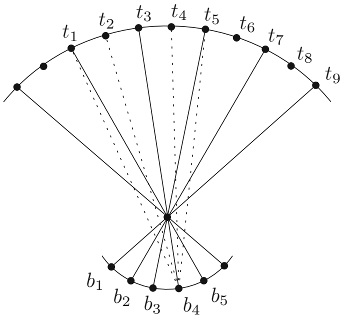
Here is an example of a typical step in the algorithm, shown in Fig. 6. Suppose some configuration includes points t_1,t_2,b_2,b_1, suppose that $D[1,1]=D[2,2]=\{2\}$, $D[1,2]=\{2,3,\infty\}$ and that $D[2,1]=\{1,2,3,\infty\}$. Then, using Fact 3.1, we know that $|t_1b_2|+|t_2b_1|>|t_1b_1|+|t_2b_2|$. As the right-hand side equals $2d_2$ and the maximum possible length of t_1b_2 is d_2, we can deduce that $|t_2b_1|>d_2$ and so we may update the configuration via $D[2,1]:=\{x\in D[2,1] \mid x<2\}=\{1\}$.

The program uses Facts 3.1–3.4 in ways analogous to the above example. Whenever one of the facts is applicable, we use it to reduce the size of one set D in the configuration. We use Fact 3.4 only when a_1,a_i,a_j lie in the top interval and a_k,a_l,a_n lie in the bottom, or vice versa.



	$i=1$	2	3	4	5	6
$j=1$	∞	∞	∞	3	2	1
$j=2$	∞	∞	3	2	1	1
$j=3$	∞	3	2	1	1	2
$j=4$	3	2	1	1	2	3
$j=5$	2	1	1	2	3	∞
$j=6$	1	1	2	3	∞	∞

Fig. 4 *Left*: an odd regular polygon, with a *top* and *bottom* interval. *Right*: the corresponding values of D, where entry x in column i, row j indicates $D[i,j] = \{x\}$. One level is illustrated on the *left* and *circled* on the *right*



	$i=1$	2	3	4	5	6	7	8	9
$j=1$	∞	∞	∞	∞	∞	∞	2	2	1
$j=2$	∞	∞	∞	∞	2	2	1	2	2
$j=3$	∞	∞	2	2	1	2	2	∞	∞
$j=4$	2	2	1	2	2	∞	∞	∞	∞
$j=5$	1	2	2	∞	∞	∞	∞	∞	∞

Fig. 5 *Left*: an illustration of Vesztergombi's construction with $m_2 = \frac{4}{3}n - O(1)$. Some diagonals of lengths d_1 and d_2 are shown (*solid and dotted*, respectively). *Right*: the corresponding configuration; again, entry x in column i, row j indicates $D[i,j] = \{x\}$

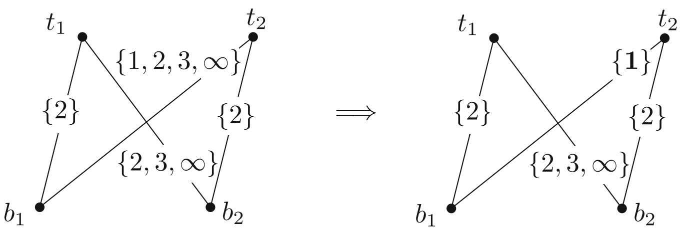


Fig. 6 A typical step of the algorithm, using Fact 3.1

Our algorithm also makes use of another easy observation. In any instance S, it cannot be true that both $d_1 + d_3 > d_2 + d_2$ and $d_1 + d_3 < d_2 + d_2$. Hence, using Fact 3.1, a quadruple t, t', b', b (in that cyclic order) with $|tb| = |t'b'| = d_2, |tb'| = d_1, |t'b| = d_3$ cannot co-exist with another quadruple $\hat{t}, \hat{t}', \hat{b}', \hat{b}$ with $|\hat{t}\hat{b}| = d_1, |\hat{t}'\hat{b}'| = d_3, |\hat{t}\hat{b}'| = |\hat{t}'\hat{b}| = d_2$. More generally, given a configuration, we can deduce from any i, j, i', j' with each $D[i,j], D[i,j'], D[i',j], D[i',j']$ singletons other than $\{\infty\}$ that an inequality of the form $d_w + d_x > d_y + d_z$ is true; in testing a configuration for validity, our program will reject any configuration where a contradiction arises from the set of all such pairwise inequalities. This is done by testing the associated digraph of $\binom{k+1}{2}$ pairs for acyclicity. (We also include arcs of the form $d_x + d_y > d_x + d_z$ whenever $y < z$.)

In some situations none of these facts is applicable; say, for example, if each $D[i,j]$ is equal to $\{1,2,\infty\}$, we cannot conclude any further information. In this case, we use an approach that is similar to recursion or *branch-and-bound* in this situation, which works as follows. Find some i,j with $|D[i,j]| > 1$, and let X denote $D[i,j]$. We then replace this configuration with two new configurations: Each of the new ones is almost identical to the original, except that in one we take $D[i,j] = \min_{x\in X} x$ and in the other we take $D[i,j] = X\backslash\{\min_{x\in X} x\}$. In a little more detail, while we are examining the levels from 1 to L, we only perform branching on diagonals in levels 1 to L, (i.e., only when $1 \le j - i \le L$) and any other nonsingleton $D[i,j]$ does not entail branching. This was faster in practice than branching on every $D[i,j]$.

4.4 Initializing and Growing Configurations

Recall that our theorems are all of the following form, for a set T of positive integers and some real α:

$$\sum_{t\in T} m_t \le \alpha n + O(1). \qquad (\spadesuit)$$

We call a *target distance* any distance d_t with $t \in T$. We use k to represent the largest number in T.

We begin this detailed section by explaining why it suffices to examine configurations of bounded size to bound the number of target distances in L consecutive levels. The key tool is Fact 3.3. Namely, suppose t_0b_1 is any diagonal in level 1 with length $|t_0b_1| \ge d_k$, and consider any top-k distance diagonal e in levels $1,\dots,L$. If e crosses t_0b_1, then t_0 (resp., b_1) is within L steps along the boundary from an endpoint of e (resp., the other endpoint of e). If e and t_0b_1 don't cross, one endpoint of e is at most $2k$ steps from t_0 or b_1 by Fact 3.3, and the other endpoint of e is at most $2k+L$ points away from the other of t_0 or b_1. Summarizing, in either case, e has one endpoint in the interval I_t consisting of vertices at most $2k+L$ steps from t_0, and e's other endpoint lies in the interval I_b consisting of vertices at most $2k+L$ steps from b_1; and this holds for all top-k distance diagonals e in levels $1,\dots,L$.

Our program makes valid deductions whenever these intervals are disjoint, which is false only when t_0 and b_1 are within $2(2k+L)$ steps of one another on the boundary. Set $\ell = 2(2k+L)$ and define a *special diagonal* to be one with length $\ge d_k$ and at most ℓ vertices between its endpoints. Recall that $|t_0b_1| \ge d_k$, so the program's deductions are valid unless there was a special diagonal. This explains the choice of $16 = 2(2\cdot 2+4)$ in Proposition 4.1 and justifies our general approach.

In the rest of this section, we explain some of the implementation details. The program begins working with a configuration consisting of a single diagonal t_0b_1 of length $\geq d_k$, and we assume without loss of generality that there are no diagonals $t_i b_{i+1}$ such that $i < 0$ and $|t_i b_{i+1}| \geq d_k$. Thus, the top and bottom intervals begins as the singleton sets $\{t_0\}, \{b_1\}$.

We will now enlarge these configurations. Reviewing our proof strategy, the program must enumerate all possible configurations such that level 1 has more than α diagonals of a target length, *and* levels 1 and 2 together have more than 2α, etc., with the hope being that once the number of levels is high enough, we find that no such configurations exist, since this would give a result like Lemma 4.2.

Note that, by our choice of t_0 and b_1, which normalize our indices, in any convex point set, all level-1 diagonals of the target distances are of the form $t_i b_{i+1}$ for $i > 1$, and by Fact 3.3, they also satisfy $i \leq 2k-2$. So crucially, their possible positions are confined to an interval of bounded size. We now determine which of these diagonals have target lengths by *exhaustive guessing*, a term that simply means trying all possibilities. In detail, first, exhaustively guess the smallest $i > 0$ for which $t_i b_{i+1}$ is a target distance, then the second-smallest, etc. When the top and bottom intervals are enlarged, each new $D[i,j]$ is set to $\{1,\ldots,k,\infty\}$ by default, meaning that no assumptions are made on the distance. When i is guessed as a minimal new level-1 diagonal for which $t_i b_{i+1}$ is a target distance, rather than the defaults, we set $D[i,i+1] = T$ and $D[i',i'+1] := \{1,\ldots,k,\infty\} \backslash T$ for all new $i' < i$.

After each new diagonal is added, we reapply Facts 3.1–3.4 in order to make additional deductions and eliminate any impossible configuration; and we split any nonsingleton sets D in the first level, as described earlier.

After this exhaustive guessing, we have collected all possible configurations. We keep only those for which level 1 has more than α diagonals of the target lengths. If any exist, we grow them in all possible ways to 2-level configurations, using exhaustive guessing like that explained above, except that we expand "to the left" before expanding "to the right" (for level 1, only rightward expansion was needed due to our choice of t_0 and b_1). Again, we prune those that have no more than 2α target distance in the first two levels.

We repeat the process described in the previous paragraph over and over, increasing the number of levels by 1 each time. If the program terminates eventually, it implies a result of the form like Lemma 4.2 and consequently that (♠) holds for this choice of T and α. We give a high-level review of the algorithm in Fig. 7.

5 Results: Proof of Theorem 4

Each row in Table 1 corresponds to an execution of our program that terminated. In other words, each execution establishes that an analogue of Lemma 4.2 holds, and we consequently deduce Theorem 4 using reasoning as in the proof of Proposition 4.1. Each line proves

$$\sum_{t \in T} m_t \leq \alpha n \text{ for } n > C(k, 2(2k+L))/(\alpha - 1), \tag{♣}$$

- Initialize a configuration with intervals $\{t_0\},\{b_1\}$ and $D[0,1]$ set to T (all target distances)
- For $L = 1, 2, \ldots$
  - Extend the configurations by exhaustively guessing all diagonals of target lengths in level L, extending leftwards first if $L > 1$, and then rightwards in all cases.
  - Keep only configurations with more than αL target distances in levels $1, \ldots, L$.
  - Stop if no configurations remain.
- Upon extending a configuration, **check** it:
  - Use Facts 3.1–3.4 to perform deductions.
  - Check that distance pairs are consistent.
  - If $|D[i,j]| > 1$ for some diagonal $t_i b_j$ in one of the first L levels, partition it into two configurations and **check** both (recursively).

Fig. 7 Sketch of the algorithm

Table 1 The terminating executions of our program, each one proving (♣) for that α and T. *Tight* means convex point sets are known with $\sum_{t \in T} m_t = \alpha n - O(1)$, and *abstractly tight* means some periodic configuration has $\sum_{t \in T} m_t = \alpha n$, but we could not realize it convexly in the plane

T	α	L	Time (s)	Tightness of result
$\{1,2\}$	2	2	<1	Tight (odd regular)
$\{2\}$	4/3	4	<1	Tight [7]
$\{1,2,3\}$	3	3	<1	Tight (odd regular)
$\{3\}$	3/2	9	5	Abstractly tight, Fig. 8
$\{2,3\}$	9/4	6	1	Abstractly tight, Fig. 9
$\{1,3\}$	2	4	<1	Tight (odd regular)
$\{1,2,3,4\}$	4	3	68	Tight (odd regular)
$\{4\}$	13/8	27	50,890	Unknown

where k is the largest element of T, and C is the constant from Lemma 1.5. Note that the first two lines of Table 1 correspond to results that were already known. The running times are from a computer with a 2-GHz processor. The program was written in Java and is available on SourceForge.[2] For $T = \{1,2,3,4,5\}$ or $T = \{5\}$, the program ran out of memory before obtaining any reasonable result.

6 Abstract Tightness

Our computer program can also generate tight examples. In Fig. 8, we show two periodic configurations with $m_3 = \frac{3}{2}n$ with periods of six and eight levels, respectively. (No other example has period lower than 14.) We were not able to embed these examples as convex point sets in the plane, and at the same time we

[2] http://sourceforge.net/projects/convexdistances/.

∞	∞	∞	∞	∞	∞	∞	∞	3	3	1	2	1	3	3
∞	∞	∞	∞	∞	∞	3	3	2	3	2	3	3	∞	∞
∞	∞	∞	∞	3	3	1	2	1	3	3	∞	∞	∞	∞
∞	∞	3	3	2	3	2	3	3	∞	∞	∞	∞	∞	∞
3	3	1	2	1	3	3	∞	∞	∞	∞	∞	∞	∞	∞

∞	∞	∞	∞	∞	∞	3	3	3
∞	∞	∞	∞	3	3	1	2	3
∞	∞	∞	3	2	3	2	3	∞
∞	∞	3	2	1	3	3	∞	∞
∞	∞	3	3	3	∞	∞	∞	∞
3	3	1	2	3	∞	∞	∞	∞

Fig. 8 Two unrealized periodic configurations with $m_3 = \frac{3}{2}n$. *Rows* and *columns* are two intervals of vertices, and entry i (resp., ∞) means distance d_i (resp., $< d_3$)

∞	∞	∞	3	2	2	2	3	∞
∞	∞	3	2	1	2	3	∞	∞
∞	3	2	2	2	3	∞	∞	∞

Fig. 9 An unrealized periodic configurations with $m_2 + m_3 = \frac{9}{4}n$

did not disprove that they were embeddable. Based on our attempts, it seems like there is no simple periodic embedding respecting the natural symmetries of the distance configurations. A disproof of realizability could be used in the program to get stronger results. For $m_2 + m_3 = \frac{9}{4}n$, we also have an abstractly tight periodic example that we could not realize (Fig. 9).

7 Future Directions

Our program is essentially a depth-first search; each configuration examined by the program has a unique "parent" configuration from which it was grown. Thus, it would be possible to rewrite the program so as to use a smaller amount of memory and thereby possibly obtain results with smaller α or larger k; and a distributed implementation should also be straightforward.

It would be good to come up with constructions exhibiting better lower bounds. For example, no construction is known where m_3/n is asymptotically greater than 4/3.

Our approach constitutes an abstract generalization of the original problem of bounding sums of the m_i's in convex point sets. Vesztergombi [7] considered an abstraction as well, using only a subset of the facts we applied here. Can Conjecture 1.1 of Erdős and Moser be violated in either of these abstractions?

Finally, can the functions C, C' in Lemmas 1.5 and 1.6 be improved?

Acknowledgements We thank the referees for useful feedback, and K. Vesztergombi for helpful discussions. EPFL, Lausanne, Switzerland. We gratefully acknowledge support from the Swiss National Science Foundation (Grant No. 200021-125287/1) and an NSERC Post-Doctoral Fellowship.

References

1. E. Altman, On a problem of P. Erdős. Am. Math. Mon. **70**(2), 148–157 (1963)
2. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, New York, 2005)
3. H. Edelsbrunner, P. Hajnal, A lower bound on the number of unit distances between the vertices of a convex polygon. J. Comb. Theor. A **56**(2), 312–316 (1991)
4. P. Erdős, L. Lovász, K. Vesztergombi, On the graph of large distances. Discrete Comput. Geom. **4**, 541–549 (1989)
5. H. Hopf, E. Pannwitz, Aufgabe Nr. 167. Jahresbericht Deutsch. Math.-Verein. **43**, 114 (1934)
6. K. Vesztergombi, On large distances in planar sets. Discrete Math. **67**, 191–198 (1987)
7. K. Vesztergombi, On the distribution of distances in finite sets in the plane. Discrete Math. **57**, 129–145 (1985)

Coloring Distance Graphs and Graphs of Diameters

Andrei M. Raigorodskii

Abstract In this chapter, we discuss two classical problems lying on the edge of graph theory and combinatorial geometry. The first problem is due to E. Nelson. It consists of coloring metric spaces in such a way that pairs of points at some prescribed distances receive different colors. The second problem is attributed to K. Borsuk and involves finding the minimum number of parts of smaller diameter into which an arbitrary bounded nonsingleton point set in a metric space can be partitioned. Both problems are easily translated into the language of graph theory, provided we consider, instead of the whole space, any (finite) distance graph or any (finite) graph of diameters. During the last decades, a huge number of ideas have been proposed for solving both problems, and many results in both directions have been obtained. In the survey below, we try to give an entire picture of this beautiful area of geometric combinatorics.

1 Basic Definitions and Motivations

The main objects of this chapter are *distance graphs* and *graphs of diameters*. In principle, one may define them for any metric space. However, we start here by defining them only in the case of the Euclidean space $\mathbb{R}^n$. More general cases will be discussed a bit later.

A.M. Raigorodskii (✉)
Moscow State University, Mechanics and Mathematics Faculty,
Department of Mathematical Statistics and Random Processes, GSP-1, 1
Leninskiye Gory, Main Building, 119991 Moscow, Russia

Moscow Institute of Physics and Technology, Faculty of Innovations
and High Technology, Department of Data Analysis, Moscow, Russia

Yandex Research Laboratories, Moscow, Russia
e-mail: raigorodsky@yandex-team.ru

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_23,

So consider an arbitrary graph $G = (V, E)$ such that $V \subset \mathbb{R}^n$, $|V| < \infty$, and

$$E \subseteq \{\{\mathbf{x}, \mathbf{y}\} : |\mathbf{x} - \mathbf{y}| = 1\}, \qquad (1)$$

where by $|\mathbf{x} - \mathbf{y}|$ we denote the standard Euclidean distance. Any such graph is called a *(finite) distance graph*. Here it is important that the set of edges of a distance graph is not necessarily formed by *all* pairs of vertices $\mathbf{x}, \mathbf{y}$ satisfying the distance condition $|\mathbf{x} - \mathbf{y}| = 1$ [cf. (1)]: we assume that some "possible" edges are not drawn. However, if in expression (1) we substitute the sign "$\subseteq$" by the sign "$=$" (which is somehow natural), then the corresponding graph G is called the *complete distance graph*.

There exists a great literature concerning distance graphs (see, e.g., the book [16]). In this chapter, we are mainly interested in coloring them. More precisely, we study the *chromatic numbers* of distance graphs, where by the "chromatic number" $\chi(G)$ of a graph G we mean, as usual, the minimum number of colors needed to color all the vertices of G so that any two adjacent vertices receive different colors.

A basic motivation for us is given by a famous problem in combinatorial geometry. This problem consists of finding the chromatic number $\chi(\mathbb{R}^n)$ *of the Euclidean space*, which is the smallest χ such that one can color all the points in $\mathbb{R}^n$ by χ colors and no two points at the distance 1 get the same color:

$$\chi(\mathbb{R}^n) = \min\left\{\chi : \mathbb{R}^n = V_1 \sqcup \ldots \sqcup V_\chi, \; \forall\, i \;\; \forall\, \mathbf{x}, \mathbf{y} \in V_i \;\; |\mathbf{x} - \mathbf{y}| \neq 1\right\}.$$

In other words, $\chi(\mathbb{R}^n) = \chi(\mathfrak{G})$, where $\mathfrak{G} = (\mathbb{R}^n, \mathfrak{E})$ and

$$\mathfrak{E} = \{\{\mathbf{x}, \mathbf{y}\} : |\mathbf{x} - \mathbf{y}| = 1\}.$$

Of course, the quantity $\chi(\mathfrak{G})$ is bounded from below by the chromatic number of any finite distance graph. A much more essential thing is that, in fact,

$$\chi(\mathfrak{G}) = \max_G \chi(G), \qquad (2)$$

where the maximum is taken over all finite distance graphs in $\mathbb{R}^n$. This result is an immediate consequence of a general theorem by Erdős and de Bruijn (see [17]): *If the chromatic number of an infinite graph G is finite, then it is attained on a finite subgraph of G.* In Sect. 2.3, we shall show, in particular, that $\chi(\mathfrak{G}) = \chi(\mathbb{R}^n) < \infty$, and so Eq. (2) will be proved.

Since we prohibit points of the same color from being at the distance 1, we also call the value 1 the *forbidden distance*. By the homogeneity of the Euclidean distance, the quantity $\chi(\mathbb{R}^n)$ will not change, provided we substitute 1 by any other forbidden distance $a > 0$.

The history of the problem of determining $\chi(\mathbb{R}^n)$ is very intriguing. The problem was proposed in 1950 by E. Nelson, although H. Hadwiger had worked on similar questions even earlier (see [48]). Some "colorful" details can be found in the book [114], and here we do not dwell on this. Many other books and surveys can be cited as well; see, e.g., [3, 16, 59, 82, 83, 115].

Now, consider an arbitrary graph $G = (V,E)$ such that $V \subset \mathbb{R}^n$, $|V| < \infty$, and

$$E = \{\{\mathbf{x},\mathbf{y}\}: \ |\mathbf{x}-\mathbf{y}| = \mathrm{diam} V\}, \tag{3}$$

where by diam$V$ we denote the diameter of V, i.e., the value

$$\mathrm{diam} V = \sup_{\mathbf{x},\mathbf{y} \in V} |\mathbf{x}-\mathbf{y}|.$$

In principle, here one may substitute "sup" by "max" since V is finite. Any graph determined by (3) is called the *graph of diameters*: We join any two points in V by an edge, if and only if they are, in some sense, "antipodal," which means that they realize the maximum distance in the set V.

In this chapter, we study only the chromatic numbers of distance graphs. This is well motivated by the famous Borsuk partition problem. Indeed, consider an arbitrary set Ω of diameter 1 in $\mathbb{R}^n$ that is not necessarily finite. By $f(\Omega)$, denote the minimum number of smaller-diameter parts needed to decompose Ω:

$$f(\Omega) = \min\left\{f: \ \Omega = \Omega_1 \sqcup \ldots \sqcup \Omega_f, \ \forall\, i \ \ \forall\, \mathbf{x},\mathbf{y} \in \Omega_i \ \ |\mathbf{x}-\mathbf{y}| < 1\right\}.$$

By $f(n)$, denote, in turn, the value $\max_{\Omega} f(\Omega)$. In other words, $f(n)$ is the lowest number of smaller-diameter parts needed to divide an arbitrary set of diameter 1 in $\mathbb{R}^n$. The Borsuk problem is just in finding $f(n)$. Notice that "diameter 1" can be replaced by "diameter a" with any $a > 0$ in all the above definitions.

For finite sets V in the Euclidean space, the chromatic numbers of the corresponding graphs of diameters are exactly equal to $f(V)$. For infinite sets, this is not the case. For example, the graph of diameters of the unit sphere $S^{n-1} \subset \mathbb{R}^n$ is obviously bipartite (it is even a matching); however, $f(S^{n-1}) = n+1$, which is essentially equivalent to the Borsuk–Ulam theorem (see [14, 76]).

The origin of the problem is in 1933, when K. Borsuk conjectured in [14] that $f(n) = n+1$. The history of this conjecture is dramatic, since the majority of specialists believed in it; nevertheless, a counterexample was constructed in 1993 (see Sect. 8.3). What is essential for us right now is that all the known counterexamples to Borsuk's conjecture are given by *finite* sets in the Euclidean spaces.

The literature on Borsuk's problem is huge; see, e.g., [13, 16, 82, 84–87].

The structure of the remaining part of the chapter will be as follows: In Sect. 2, we shall discuss various questions concerning the value $\chi(\mathbb{R}^n)$; Sect. 3 will be devoted to the chromatic numbers of more general metric spaces; in Sect. 4, we shall speak about distance graphs with a large girth/small clique number and a high chromatic number; in Sect. 5, we shall introduce a sequence of random distance graphs related to the coloring problems; in Sect. 6, we shall discuss the independence numbers of some important distance graphs; Sect. 7 will be devoted to some "conditional" results on various chromatic numbers; in Sect. 8, we shall eventually proceed to Borsuk's problem.

2 The Chromatic Numbers of $\mathbb{R}^n$

2.1 The Case $n = 2$

We start by considering $n = 2$, since of course $\chi(\mathbb{R}^1) = 2$ and almost nothing of interest can be found in the one-dimensional case (see, however, Sect. 3.4.2). One even could assume that finding $\chi(\mathbb{R}^2)$ should be not a big deal. Unexpectedly, such an assumption is completely wrong. We only know that

$$4 \leqslant \chi(\mathbb{R}^2) \leqslant 7.$$

Moreover, both estimates are very simple. They were obtained just after the problem had been stated, and to date, no one knows how to improve them.

As for the lower bound, it is attained on a distance graph, which is called *Moser's spindle* (see Fig. 1).

As for the upper bound, it is given by an explicit coloring of the plane, which is shown in Fig. 2.

Although the quantity $\chi(\mathbb{R}^2)$ has yet to be determined, many facts are known about coloring distance graphs in the Euclidean plane. For example, P. Erdős, who liked the problem very much, proposed the following question in 1976 (see [29]): *Does there exist a triangle-free distance graph in the plane whose chromatic number is at least 4?* Erdős's question is quite natural, since, on the one hand, Moser's spindle contains four triangles and, on the other hand, there do exist ordinary graphs G with any prescribed value of the chromatic number and any prescribed value of the *girth* $g(G)$ (which is the length of a shortest cycle in the graph).

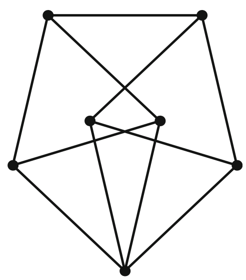


Fig. 1 Moser's spindle

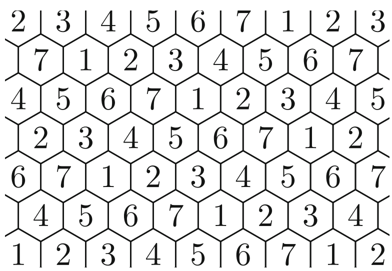


Fig. 2 Coloring of the plane

In 1979, N. Wormald succeeded in positively answering Erdős's question. He constructed a distance graph G in $\mathbb{R}^2$ with $\chi(G) = 4$ and $g(G) = 4$. His graph had 6448 vertices (see [118]). In 1996, a similar graph with only 23 vertices was discovered by O'Donnell and Hochberg (see [25]). Finally, in 2000, O'Donnell showed the existence, for any fixed k, of a distance graph with $\chi(G) = 4$ and $g(G) > k$ (see [26, 27]).

Another direction of research is generated by imposing various restrictions on the colors. One may assume that only *measurable* sets in the plane must be considered as possible colors in a coloring. In this case, the chromatic number is also called *measurable* and is denoted by $\chi_m(\mathbb{R}^2)$. K. Falconer proved in 1981 that $\chi_m(\mathbb{R}^2) \geqslant 5$ (see [37]).

Furthermore, one may assume that any color consists of pairwise disjoint sets whose boundaries are formed by finitely many Jordan arcs. With this restriction, five colors do not already suffice and one needs at least six colors. This was noticed by D.R. Woodall in 1973 (see [117]).

On the other hand, by analyzing the coloring from Fig. 2 more carefully, we see that there is some room to spare. Indeed, not only is the distance 1 avoided there, but a whole interval of distances can be forbidden with the same effect. Starting from this point, L.L. Ivanov recently studied the value $\chi(\mathbb{R}^2; [1,d])$ that is equal to the minimum number of colors needed to color the plane, so that any two points at a distance $a \in [1,d]$ receive different colors (see [51]). In particular, he proved that

$$\chi\left(\mathbb{R}^2; \left[1, \frac{\sqrt{7}}{2}\right]\right) \leqslant 7, \quad \chi\left(\mathbb{R}^2; \left[1, \sqrt{3}\right]\right) \leqslant 9, \quad \chi\left(\mathbb{R}^2; [1,2]\right) \leqslant 12,$$

$$\chi\left(\mathbb{R}^2; \left[1, \frac{\sqrt{19}}{2}\right]\right) \leqslant 13, \quad \chi\left(\mathbb{R}^2; \left[1, \frac{3\sqrt{3}}{2}\right]\right) \leqslant 16, \quad \chi\left(\mathbb{R}^2; \left[1, \frac{\sqrt{31}}{2}\right]\right) \leqslant 19,$$

$$\chi\left(\mathbb{R}^2; \left[1, \frac{\sqrt{37}}{2}\right]\right) \leqslant 21.$$

Many analogous more or less reasonable problems have been formulated and more or less solved (see [114]), but the original question is still far from being answered. We would like to conjecture that $\chi(\mathbb{R}^2) = 4$. However, it is worth noting that if our conjecture is true, then to prove it, we must find a 4-coloring of the plane with nonmeasurable sets as colors. To date, no techniques have been developed to do that.

2.2 *Other "Small" Dimensions*

For the space $\mathbb{R}^3$, the gap between upper and lower estimates of the chromatic number is much bigger than the similar gap for the plane. Now, we know that

$$6 \leqslant \chi(\mathbb{R}^3) \leqslant 15.$$

The lower bound is due to Nechushtan (see [79]) and the upper one was obtained by Coulson (see [22]) and, independently, by Radoičić and Tóth (see [81]).

For $\mathbb{R}^4$, the best-known lower bound is by 7. The first proof of this bound was given by Cantwell in [18]. The corresponding distance graph is rather complicated. Substantially simpler graphs were constructed by Ivanov (see [52]) and Kupavskii (see [61]).

The bound $\chi(\mathbb{R}^4) \leqslant 49$ was announced by Coulson in 2000, but as far as we know, no corresponding publication has appeared since then. Thus, the best published upper estimate for which we can give a reference is by Radoičić and Tóth (see [81]): $\chi(\mathbb{R}^4) \leqslant 54$.

A tremendous volume of publications concerning lower bounds for $\chi(\mathbb{R}^n)$ with $n \geqslant 5$ may be cited. Instead of doing that, we shall give a table with the current "records" for $n \leqslant 12$.

dim	1	2	3	4	5	6
$\chi \geqslant$	2	4	6, [79]	7, [18, 52, 61]	9, [18]	11, [20]
dim	7	8	9	10	11	12
$\chi \geqslant$	15, [82]	16, [68]	21, [65]	23, [65]	25, [62]	27, [61]

The measurable chromatic number can be defined not only for the plane (see Sect. 2.1), but for any space $\mathbb{R}^n$. The value $\chi_m(\mathbb{R}^n)$ has been studied by many authors, including Falconer (see [37]) and Székely (see [115] and [116]). The best lower bounds for this value are as follows:

$$\chi_m(\mathbb{R}^3) \geqslant 7,\ \chi_m(\mathbb{R}^4) \geqslant 9,\ \chi_m(\mathbb{R}^5) \geqslant 14,\ \chi_m(\mathbb{R}^6) \geqslant 20,\ \chi_m(\mathbb{R}^7) \geqslant 28,\ \chi_m(\mathbb{R}^8) \geqslant 39,$$
$$\chi_m(\mathbb{R}^9) \geqslant 54,\ \chi_m(\mathbb{R}^{10}) \geqslant 73,\ \chi_m(\mathbb{R}^{11}) \geqslant 97,\ \chi_m(\mathbb{R}^{12}) \geqslant 129, \ldots$$

All these bounds were obtained by de Oliveira Filho and Vallentin in their recent joint paper [38].

Also, the value $\chi(\mathbb{R}^3; [1,d])$ (cf. Sect. 2.1) was studied in [51]

$$\chi\left(\mathbb{R}^3; [1, 1.115]\right) \leqslant 18,\quad \chi\left(\mathbb{R}^3; [1, 1.133]\right) \leqslant 21,\quad \chi\left(\mathbb{R}^3; [1, 1.137]\right) \leqslant 23,$$
$$\chi\left(\mathbb{R}^3; [1, 1.303]\right) \leqslant 24,\quad \chi\left(\mathbb{R}^3; [1, 1.549]\right) \leqslant 27, \ldots$$

2.3 Higher Dimensions

It is very easy to show that $\chi(\mathbb{R}^n) < \infty$. Indeed, one can partition all the space into very big equal "parallel" cubes (say, of side length 2) and then divide every such cube into very small cubes (say, of side length $< \frac{1}{2\sqrt{n}}$). In any big cube, we get as many colors as we have small cubes in it. So the distance between monochromatic points in a given small cube is less than 1, and the distance between points in any two distinct small monochromatic cubes is greater than 1. Careful calculations provide here a bound $\chi(\mathbb{R}^n) \leqslant (\lceil\sqrt{n}\rceil + 1)^n$.

Much tighter bounds were obtained by Larman and Rogers in 1972 in [68]: $\chi(\mathbb{R}^n) \leqslant (3 + o(1))^n$. This result is still the best known. Even for $n \geqslant 5$, no other more specific estimates have been established.

As for the lower estimates, one can readily generalize Moser's spindle in order to get $\chi(\mathbb{R}^n) \geqslant n+2$, $n \geqslant 2$. The first improvement to this almost trivial bound was given in [68]: $\chi(\mathbb{R}^n) \geqslant \frac{\binom{n}{3}}{n} \sim \frac{n^2}{6}$. In 1978, Larman obtained a slightly further refinement (see [67]): $\chi(\mathbb{R}^n) \geqslant cn^3$, $c > 0$. A breakthrough was done in a seminal paper by Frankl and Wilson (see [43]), who succeeded in showing in 1981 that

$$\chi(\mathbb{R}^n) \geqslant \left(\frac{1+\sqrt{2}}{2} + o(1)\right)^n = (1.207\ldots + o(1))^n.$$

The current "record" here is due to this author (see [82, 88]), so that

$$(1.239\ldots + o(1))^n \leqslant \chi(\mathbb{R}^n) \leqslant (3 + o(1))^n. \tag{4}$$

In Sect. 2.2, we could see a big difference between lower bounds for the ordinary chromatic number of space and for the measurable one. It seems natural to guess that there should be a large gap between similar bounds in growing dimension as well. Unexpectedly, only the same estimates (4) are known for $\chi_m(\mathbb{R}^n)$. Of course, infinitesimals $o(1)$ are not the same, but the main exponent does not change. Moreover, there are no examples when $\chi_m(\mathbb{R}^n) \neq \chi(\mathbb{R}^n)$.

2.4 *Ideas of Lower Estimates in Small Dimensions*

Of course, the first step in getting lower estimates for $\chi(\mathbb{R}^n)$ is always the same: One should find a distance graph in $\mathbb{R}^n$ with as high a chromatic number as possible. In the plane, such a graph is just Moser's spindle, but in greater dimensions, constructions are sometimes quite nontrivial. Many of them are described in [115], and we do not want to dwell on them here.

A more interesting type of argument is suggested by a careful analysis of Moser's spindle. How do we construct it? We take two points $\mathbf{x}, \mathbf{y}$ at distance 1. Then we find two other points $\mathbf{z}, \mathbf{z}'$ such that each of them is at distance 1 from both $\mathbf{x}$ and $\mathbf{y}$. Finally, we rotate the construction in order to get the spindle. The distance graph on the vertices $\mathbf{x}, \mathbf{y}$ has chromatic number 2. Both distance graphs on the vertices $\mathbf{x}, \mathbf{y}, \mathbf{z}$ and $\mathbf{x}, \mathbf{y}, \mathbf{z}'$ (triangles) have chromatic number 3. And the whole spindle has chromatic number 4. Starting from a graph H with chromatic number 2, we come to another graph G, whose chromatic number is already 4.

An important extension of the above-described construction is as follows. Assume that we have a graph H whose vertices lie on a sphere $S_r^{n-1} \subset \mathbb{R}^n$ of some radius r (in Moser's spindle, H lies on $S_1^0 \subset \mathbb{R}^1$). Moreover, let $\chi(H) = m$. Suppose that there is a point $\mathbf{z}$ in $\mathbb{R}^{n+1}$ that is at distance 1 from every point of S_r^{n-1} (the existence of such a point depends only on the value of r). Then we immediately get a graph G in $\mathbb{R}^{n+1}$ with $\chi(G) = m+1$. Now, take the point $\mathbf{z}'$, which is symmetric to $\mathbf{z}$ and also has distance 1 from S_r^{n-1}. Rotate this construction, and the resulting spindle has chromatic number $m+2$.

The above procedure gives us a possibility to *lift* a lower bound from a dimension n to the next dimension. Kupavskii thoroughly investigated more subtle constructions, obtaining a series of similar but stronger results (see [61, 62]). We quote here only one of such results: *Let G be a distance graph whose vertices lie on a sphere $S_r^{n-1} \subset \mathbb{R}^n$ with $n \geqslant 2$ and*

$$\frac{1}{2} \leqslant r \leqslant \sqrt{\frac{1+\sqrt{3}}{2+\sqrt{3}}}, \quad r \neq \sqrt{\frac{2}{3}}.$$

Assume that $\chi(G) \geqslant m$. Then $\chi(\mathbb{R}^{n+2}) \geqslant m+4$. In other words, we can add 4 to a bound of the chromatic number by increasing the space dimension by 2.

For example, in the table from Sect. 2.2, we see the estimates

$$\chi(\mathbb{R}^9) \geqslant 21, \quad \chi(\mathbb{R}^{10}) \geqslant 23, \quad \chi(\mathbb{R}^{11}) \geqslant 25, \quad \chi(\mathbb{R}^{12}) \geqslant 27.$$

The first estimate is attained on a graph whose vertices just lie on a sphere (see [65]). The second one is due to the "spindle idea." The third one is obtained by using the quoted result of Kupavskii. The fourth bound is based on another result from Kupavskii's work [61].

2.5 Ideas of Lower Estimates in Higher Dimensions

As we know from Sect. 2.3, the first nontrivial estimate is $\chi(\mathbb{R}^n) \geqslant \frac{\binom{n}{3}}{n} \sim \frac{n^2}{6}$. It is published in the paper [68], but it was suggested by Erdős and Sós. The corresponding graph $G = (V,E)$ is as follows:

$$V = \{\mathbf{x} = (x_1,\ldots,x_n) : \ x_i \in \{0,1\},\ x_1+\cdots+x_n = 3\}, \quad E = \{\{\mathbf{x},\mathbf{y}\} : \ |\mathbf{x}-\mathbf{y}| = 2\}. \tag{5}$$

So here the forbidden distance is 2, but for the chromatic number, this does not matter (see Sect. 1). A more important thing is that the consideration of such a graph G leads us away from pure geometry and gives us the possibility of using some instruments from combinatorics and coding theory.

By the pigeon-hole principle, $\chi(G) \geqslant \frac{|V|}{\alpha(G)}$, where

$$\alpha(G) = \max\{|W| : \ W \subseteq V, \ \forall\, x,y \in V \ \ \{x,y\} \notin E\}$$

is the *independence number* of a graph G. In the case of the graph G defined in (5), $|V| = \binom{n}{3}$. As for $\alpha(G)$, it can be calculated explicitly:

$$\alpha(G) = \begin{cases} n, & n \equiv 0 \pmod 4, \\ n-1, & n \equiv 1 \pmod 4, \\ n-2, & n \equiv 2 \ \text{or}\ 3 \pmod 4. \end{cases}$$

The lower estimate for $\alpha(G)$ is given by a simple construction, and the upper estimate can be proved by induction.

Now, an idea of further improving the quadratic lower bound for the chromatic number of space is evident: Replace the numbers 3 (the quantity of units in each vector) and 2 (the forbidden distance) by some $a \in \{1,\dots,n\}$ and b with $b^2 \in \mathbb{N}$ and try to find the chromatic number of the corresponding graph $G_{a,b}$.

Unfortunately, this is not so simple. For instance, Larman's cubic bound (see [67] and Sect. 2.3) is obtained by taking $a = 5$, $b = \sqrt{6}$. The constant $c > 0$ in it is very small due to technical reasons. And until the appearance of the paper [43] of Frankl and Wilson, no one knew how to act when $a > 5$: Elementary tools became too cumbersome in that cases.

The paper [43] was a real breakthrough. The so-called linear algebra method was proposed there. Today we may cite many books and papers concerning this method; see, e.g., [7, 9, 82, 89]. Below we shall present a few starting ideas of the approach.

Consider the graph G from (5). Note that the forbidden distance 2 is the same as the forbidden scalar product 1. Let's denote the scalar product by $(\mathbf{x},\mathbf{y})$. Let $W = \{\mathbf{x}_1,\dots,\mathbf{x}_s\}$ be an *independent set* of vertices in V (by an "independent set," we mean any set whose elements are pairwise nonadjacent in the graph, so that the independence number is just the maximum cardinality of an independent set). In other words, $(\mathbf{x}_i,\mathbf{x}_j) \neq 1$ for any $i \neq j$, and our purpose is to give an upper estimate for the cardinality $|W| = s$.

To each vector $\mathbf{x} \in V$ we assign a polynomial $F_{\mathbf{x}} \in \mathbb{Z}/2\mathbb{Z}[y_1,\dots,y_n]$ defined by the expression

$$F_{\mathbf{x}}(\mathbf{y}) = F_{\mathbf{x}}(y_1,\dots,y_n) = (\mathbf{x},\mathbf{y}) = x_1y_1 + \cdots + x_ny_n.$$

Consider the polynomials $F_{\mathbf{x}_1},\dots,F_{\mathbf{x}_s}$. Consecutively substituting, in any linear combination of these polynomials, the variable $\mathbf{y}$ by $\mathbf{x}_i$, $i = 1,\dots,s$, we see that $F_{\mathbf{x}_i}(\mathbf{y}) = 3 \not\equiv 0 \pmod 2$ and $F_{\mathbf{x}_j}(\mathbf{y}) \in \{0,2\} \equiv 0 \pmod 2$ $(j \neq i)$. Therefore, the polynomials are linearly independent over $\mathbb{Z}/2\mathbb{Z}$. But all such polynomials are generated by $y_1,\dots,y_n$, so that the quantity $s = |W|$ of them does not exceed n and we get the bound $\alpha(G) \leqslant n$. It is only a bit different from the exact value of the independence number.

By using this powerful idea, one can prove the estimate $\alpha(G) \leqslant \binom{n}{2} + \binom{n}{1} + 1$ for Larman's graph with $a = 5$, $b = \sqrt{6}$. One should only use other polynomials. Indeed, let $F_{\mathbf{x}} \in \mathbb{Z}/3\mathbb{Z}[y_1, \ldots, y_n]$ be defined by the formula

$$F_{\mathbf{x}}(\mathbf{y}) = F_{\mathbf{x}}(y_1, \ldots, y_n) = (\mathbf{x}, \mathbf{y})((\mathbf{x}, \mathbf{y}) - 1).$$

Open the brackets in the last expression and replace any y_i^2 by y_i. We get a polynomial $\tilde{F}_{\mathbf{x}} \in \mathbb{Z}/3\mathbb{Z}[y_1, \ldots, y_n]$, which is multilinear of degree 2. Now, take an arbitrary independent set $W = \{\mathbf{x}_1, \ldots, \mathbf{x}_s\}$, i.e., $|\mathbf{x}_i - \mathbf{x}_j| \neq \sqrt{6}$ (or, equivalently, $(\mathbf{x}_i, \mathbf{x}_j) \neq 2$) for any $i \neq j$. As before, we see that $\tilde{F}_{\mathbf{x}_i}(\mathbf{x}_i) = 20 \not\equiv 0 \pmod 3$ and $\tilde{F}_{\mathbf{x}_j}(\mathbf{x}_i) \in \{0, 6, 12\} \equiv 0 \pmod 3$ $(j \neq i)$. Consequently, the polynomials are again linearly independent over $\mathbb{Z}/3\mathbb{Z}$. But all such polynomials are generated by $y_i y_j, y_i, 1$, and we are done.

Note that the just-proved estimate is much better than the original estimate by Larman. Note also that the important ingredients in the proof were as follows: (1) For any (0,1)-vector $\mathbf{y} = (y_1, \ldots, y_n)$ and for any $k \in \mathbb{N}$, $y_i^k = y_i$; (2) the forbidden scalar product is congruent modulo some p to the maximum scalar product [which is equal to $(\mathbf{x}, \mathbf{x})$, since the number a of units in every $\mathbf{x}$ is fixed]; (3) among possible scalar products between vectors from V, there are no other values congruent modulo p to the forbidden scalar product and to the maximum one; (4) p is prime.

Eventually, we are led to the statement of the Frankl–Wilson theorem, which now looks quite natural: *Let a be an arbitrary natural number not exceeding n. Let p be the minimum prime such that $a - 2p < 0$. Let the forbidden scalar product be equal to $a - p$. Then $\alpha(G_{a,b}) \leqslant \sum_{i=0}^{p-1} \binom{n}{i}$.*

From this theorem, we immediately get the estimate

$$\chi(\mathbb{R}^n) \geqslant \max_a \frac{\binom{n}{a}}{\sum_{i=0}^{p-1} \binom{n}{i}}.$$

Some simple but tedious optimization shows that the best a is asymptotically equal to $\left(\frac{2-\sqrt{2}}{2}\right) n$, entailing the bound by $(1.207\ldots + o(1))^n$ (see Sect. 2.3). For more details, the reader can study [89].

Note that, instead of primes, it is possible to take prime powers. However, replacing prime powers by anything else is a great problem (see, e.g., [42, 82, 89]).

The best-known estimate (4) is obtained by replacing (0,1)-vectors by $(-1, 0, 1)$-vectors. No other types of vectors give any improvement. We shall discuss this problem in Sect. 6.

2.6 Ideas of Upper Estimates

First, let's review some terminology from the geometry of numbers and theory of packings (see [21] and [46]).

By a *lattice* in the space $\mathbb{R}^n$, we mean the integer span of any n linearly independent vectors. Any subset Γ of a lattice Λ, which also forms a lattice, is called its *sublattice*. Lattices are abelian groups in $\mathbb{R}^n$. The *fundamental cell* of a lattice $\Lambda \subset \mathbb{R}^n$ is defined as $\mathbb{R}^n/\Lambda$. The *determinant* $\det \Lambda$ of a lattice Λ is the volume of its fundamental cell. If Λ is a lattice and Γ is its sublattice, then $|\Lambda/\Gamma| = \det \Gamma / \det \Lambda$. The *Voronoi region* $\mathcal{V}_{\mathbf{a}}$ for a point $\mathbf{a} \in \Lambda$ is defined as follows:

$$\mathcal{V}_{\mathbf{a}} = \{\, \mathbf{x} \in \mathbb{R}^n : \, |\, \mathbf{x} - \mathbf{a}| \leqslant |\, \mathbf{x} - \mathbf{b}| \ \forall \ \mathbf{b} \in \Lambda \}.$$

The set of all Voronoi regions assigned to a lattice forms a *packing* of the space by congruent polytopes, which means that those polytopes completely fill the space and intersect only by boundaries.

In these terms, the hexagonal tiling of the plane from Fig. 2 is the packing by Voronoi regions for the lattice spanned by the vectors $(1,0)$ and $\left(\frac{1}{2}, \frac{\sqrt{3}}{2}\right)$.

A general method is as follows. Take some lattice Λ and its sublattice Γ. Consider the Voronoi partition of the space corresponding to Λ. Denote by k the quantity $|\Lambda/\Gamma|$. Thus, we have k classes of points in Λ and, therefore, k classes of Voronoi regions. Color any region from the ith class into the ith color. If the diameter of a region is smaller than 1 and the distance between any two monochromatic regions is greater than 1, then we can say that $\chi(\mathbb{R}^n) \leqslant k$.

All the known upper estimates for the chromatic number of space [including those for $\chi(\mathbb{R}^n; [1,d])$] were obtained using this method. Nothing better is known!

Coming back to a discussion in Sect. 2.1, one can show that the "lattice–sublattice" techniques will never give a bound for $\chi(\mathbb{R}^2)$ better than $\chi(\mathbb{R}^2) \leqslant 7$. For $\mathbb{R}^3$, the idea is also exhausted: It is not possible to make estimate by 14 when coloring Voronoi regions. In arbitrary dimensions, the similar limitation is $2^{n+1} - 1$ (see [22]). So for $n \geqslant 4$, we have only such estimates that could be improved with the help of the existing approach.

3 The Chromatic Numbers of Some Other Spaces

3.1 A General Definition

Let (X, ρ) be an arbitrary metric space. Let $\mathcal{A} \subseteq \mathbb{R}_+$ be any set of positive reals. We emphasize that $\mathcal{A}$ may be infinite. By the *chromatic number of the space* (X, ρ) *with the set* $\mathcal{A}$ *of forbidden distances*, we mean the minimum quantity $\chi((X,\rho);\mathcal{A})$ of colors needed to color all the points in X so that any two points at a distance from $\mathcal{A}$ have different colors:

$$\chi((X,\rho);\mathcal{A}) = \min\left\{\chi : \ X = V_1 \sqcup \ldots \sqcup V_{\chi}, \ \forall\, i \ \ \forall\, x,y \in V_i \ \ \rho(x,y) \notin \mathcal{A}\right\}.$$

Denote by l_p, $p \geqslant 1$, the following metric on vectors $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ in any space X, which can be considered as the product of some spaces $X_1, \ldots, X_n$:

$$l_p(\mathbf{x}, \mathbf{y}) = \begin{cases} \sqrt[p]{|x_1 - y_1|^p + \cdots + |x_n - y_n|^p}, & p < \infty, \\ \max_i |x_i - y_i|, & p = \infty. \end{cases}$$

In this notation,

$$\chi(\mathbb{R}^n) = \chi((\mathbb{R}^n, l_2); \{1\}), \quad \chi(\mathbb{R}^n; [1,d]) = \chi((\mathbb{R}^n, l_2); [1,d]).$$

So when we omit a metric, we assume that it is the Euclidean one, and when we omit a set of forbidden distances, we assume that it consists only of the *unit*. It is worth noting here that, in an arbitrary metric space, avoiding the unit distance is not necessarily the same as avoiding another single distance.

The notion of a *distance graph* does not change essentially. When $\mathcal{A}$ is finite, there is still no need to consider infinite graphs, provided the chromatic number of a space under consideration is finite (cf. [17] and Sect. 1). However, in the case of $|\mathcal{A}| = \infty$, infinite graphs might be useful.

In Sect. 3.2, we shall discuss the chromatic numbers of rational spaces $\mathbb{Q}^n$ with the Euclidean metric and one forbidden distance; in Sect. 3.3, we shall proceed to the spaces $(\mathbb{R}^n, l_q)$, $(\mathbb{Q}^n, l_q)$ with arbitrary q and one forbidden distance; Sect. 3.4 will be devoted to multiple forbidden distances in various cases; in Sect. 3.5, we shall consider spheres in the Euclidean spaces.

3.2 The Chromatic Numbers of the Spaces $(\mathbb{Q}^n, l_2)$

The quantity $\chi(\mathbb{Q}^n) = \chi((\mathbb{Q}^n; l_2); \{1\})$ was introduced in 1976 by M. Benda and M. Perles. Their joint paper remained unpublished until 2000, when it finally appeared in *Geombinatorics* (see [11]).

Clearly, the forbidden distance 1 can be substituted here with any other rational number, but irrational forbidden distances (which must, of course, be quadratic irrationalities) may change the value of the chromatic number. Below we shall discuss only the unit distance case, referring the reader to [11].

The most interesting thing is that the value of $\chi(\mathbb{Q}^n)$ is calculated for $n \leqslant 4$:

$$\chi(\mathbb{Q}^1) = \chi(\mathbb{Q}^2) = \chi(\mathbb{Q}^3) = 2, \quad \chi(\mathbb{Q}^4) = 4.$$

As usual, the lower estimates are given by some explicit constructions (finite distance graphs). However, the upper bounds apparently do not use any lattice–sublattice technology. They are based on purely arithmetic arguments. For example, the main idea of proving the inequality $\chi(\mathbb{Q}^2) \leqslant 2$ is in the fact that if $\left(\frac{p_1}{q_1}\right)^2 + \left(\frac{p_2}{q_2}\right)^2 = 1$ and both fractions are irreducible, then q_1 and q_2 must be odd and, among p_1 and p_2, exactly one number is even (see [11, 53, 117]).

Chilakamarri conjectured in [19] that $\chi(\mathbb{Q}^5) = 8$. In 2008, Cibulka proved the bound $\chi(\mathbb{Q}^5) \geqslant 8$ (see [20]). It only remains to show that this bound is tight. However, no good estimates are known.

An up-to-date table of lower bounds is given below:

dim	1	2	3	4	5	6	7	8
$\chi \geqslant$	2	2, [117]	2, [53]	4, [11]	8, [20]	10, [75]	15, [20]	16, [75]

In growing dimension, no other idea for getting an upper estimate is known than the one described in Sect. 2.6. Since $\mathbb{Q}^n \subset \mathbb{R}^n$, we certainly have $\chi(\mathbb{Q}^n) \leqslant (3+o(1))^n$. This result is still the best known.

As for lower estimates, one uses (0,1)-graphs and $(-1,0,1)$-graphs introduced in Sect. 2.5. However, there is an additional problem. It is in the fact that any forbidden distance generating the edge set of a graph *must* be rational. In the Frankl–Wilson theorem, the forbidden scalar product equals $a - p$ (see Sect. 2.5), which is equivalent to the forbidden distance $\sqrt{2p}$. Since p is allowed to be a prime power, we may take $p = 2^{2k+1}$ in order to get $\sqrt{2p} \in \mathbb{Q}$, but this is the only possibility for getting that. So the corresponding optimization procedure is more complicated and more restricted, which leads to the best-known estimate $\chi(\mathbb{Q}^n) \geqslant (1.173\ldots+o(1))^n$ (see [82]).

Notice that the chromatic numbers of algebraic extensions of $\mathbb{Q}^n$ have been deeply studied as well. An almost exhaustive list of such results can be found in [54] (see also [114]).

To complete the section, we would like to emphasize the following facts. In small dimensions, the situation with the rational spaces is much better than the situation with the real ones. However, in asymptotics, the picture is opposite: The gap between the upper and lower estimates of the quantity $\chi(\mathbb{Q}^n)$ is even larger than the same gap in the case of $\mathbb{R}^n$.

3.3 The Chromatic Numbers of the Spaces $(\mathbb{R}^n, l_q)$, $(\mathbb{Q}^n, l_q)$

In this section, we consider colorings of the spaces $(\mathbb{R}^n, l_q)$, $(\mathbb{Q}^n, l_q)$ with 1 as the only forbidden distance. We also do not dwell on small dimensions, since too many almost trivial facts can be proved here, but there is no interest in listing them, and no deeper results are known.

When $n \to \infty$, studying l_q-metrics is much more interesting. Nevertheless, here there are some very simple results as well. The first one of them tells us that

$$\chi((\mathbb{Q}^n, l_\infty); \{1\}) = \chi((\mathbb{R}^n, l_\infty); \{1\}) = 2^n.$$

The lower estimate is given by the set of vertices of the cube, and the upper estimate is obtained by factorizing $\mathbb{R}^n$ modulo $\mathbb{Z}^n$ (see [117]).

The second simple result follows immediately from the Frankl–Wilson construction (see Sect. 2.5). Indeed, taking the forbidden scalar product $a - p$ [which, for (0,1)-vectors with a units each, is just the cardinality of intersection of their sets of units], we uniquely determine the corresponding forbidden distance $\sqrt[q]{2p}$. So just the same estimate by $(1.207\ldots + o(1))^n$ applies for *all* $q \geqslant 1$. In view of this, it is very strange to see, in the paper [57], a long proof of the bound $\chi((\mathbb{R}^n, l_1); \{1\}) \geqslant (1.067\ldots + o(1))^n$, which is apparently much worse than the trivial one.

Especially for the case of a l_1-metric, it is possible to use $(-1,0,1)$-constructions, and the best lower bound here is $\chi((\mathbb{R}^n, l_1); \{1\}) \geqslant (1.365\ldots + o(1))^n$ (see [90]). The two cases of l_1 and l_2 are the only ones where $(-1,0,1)$-graphs are better than (0,1)-configurations. So now we have the following table of lower estimates:

q	1	2	∞	Others
$\chi \geqslant$	$(1.365 + o(1))^n$, [90]	$(1.239\ldots + o(1))^n$, [88]	2^n	$(1.207\ldots + o(1))^n$, [43]

The first upper estimates for $\chi((\mathbb{R}^n, l_q); \{1\})$ were obtained in [57], but all of them are now surpassed. In [63], Kupavskii proved a series of results that we present below:

1. For any q, $\chi((\mathbb{R}^n, l_q); \{1\}) \leqslant \frac{(\ln n + \ln\ln n + \ln 4 + 1 + o(1))}{\ln\sqrt{2}} \cdot 4^n$. Actually, the same inequality holds true for an arbitrary norm in $\mathbb{R}^n$ induced by a centrally symmetric convex body.
2. For all $q > 2$, there exist c_q, $c_q < 1$, $c_q \to 0$ as $q \to \infty$, and δ_n, $\delta_n \to 0$ as $n \to \infty$, such that $\chi((\mathbb{R}^n, l_q); \{1\}) \leqslant 2^{(1+c_q+\delta_n)n}$.

So the estimate by $(4 + o(1))^n$ applies for any norm, and the estimate by $(2 + \gamma_{n,q})^n$, $\gamma_{n,q} \to 0$ as $n, q \to \infty$, is relevant for large indices q. Note that this estimate is very natural and somewhat close to an optimal one, since $\chi((\mathbb{R}^n, l_\infty); \{1\}) = 2^n$.

Now we proceed to the rational spaces. We already know that $\chi((\mathbb{Q}^n, l_\infty); \{1\}) = 2^n$ and $\chi((\mathbb{Q}^n, l_2); \{1\}) \geqslant (1.173.. + o(1))^n$. It is also clear that $\chi((\mathbb{Q}^n, l_1); \{1\}) \geqslant (1.365\ldots + o(1))^n$. Indeed, in a l_1-metric, the distances between rational vectors are rational and $(-1,0,1)$-vectors do certainly belong to $\mathbb{Q}^n$. For other metrics, the situation is quite bad. *Nothing* is known about $q \notin \mathbb{N}$. As for $q \in \mathbb{N}$, there was a result by this author (see [90]): *If* $q \geqslant 3$, *then for* $u = 2^{-q-1}$,

$$\chi((\mathbb{Q}^n, l_q); \{1\}) \geqslant \left(\frac{\left(\frac{u}{2}\right)^{\frac{u}{2}} \left(1 - \frac{u}{2}\right)^{1-\frac{u}{2}}}{u^u (1-u)^{1-u}} + o(1) \right)^n.$$

Unfortunately, the right-hand side of the last inequality tends to $(1+o(1))^n$ as q tends to infinity.

The only upper estimate for $\chi((\mathbb{Q}^n, l_q); \{1\})$ is by $\chi((\mathbb{R}^n, l_q); \{1\})$.

3.4 Multiple Forbidden Distances

3.4.1 Finite Sets of Forbidden Distances

Erdős knew (see [115]) that

$$c_1 k\sqrt{\ln k} \leqslant \max_{\mathcal{A}:\ |\mathcal{A}|=k} \chi((\mathbb{R}^2, l_2); \mathcal{A}) \leqslant c_2 k^2, \qquad c_1, c_2 > 0.$$

During the last decade, many papers have appeared around similar subjects. Now, there are, essentially, four types of values under study:

$$\overline{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); k) = \max_{\mathcal{A}\subset\mathbb{R}_+:\ |\mathcal{A}|=k} \chi((\mathbb{R}^n, l_2); \mathcal{A}), \quad \overline{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); k) = \max_{\mathcal{A}\subset\mathbb{Q}_+:\ |\mathcal{A}|=k} \chi((\mathbb{R}^n, l_2); \mathcal{A}),$$

$$\overline{\chi}_{\mathbb{R}}((\mathbb{Q}^n, l_2); k) = \max_{\mathcal{A}\subset\mathbb{R}_+:\ |\mathcal{A}|=k} \chi((\mathbb{Q}^n, l_2); \mathcal{A}), \quad \overline{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); k) = \max_{\mathcal{A}\subset\mathbb{Q}_+:\ |\mathcal{A}|=k} \chi((\mathbb{Q}^n, l_2); \mathcal{A}).$$

For the quantities with "rational" indices, almost nothing is known. Of course,

$$\begin{aligned}
\overline{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); k) &\geqslant \overline{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); 1) = \chi(\mathbb{R}^n) \geqslant (1.239\ldots + o(1))^n, \\
\overline{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); k) &\geqslant \overline{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); 1) = \chi(\mathbb{Q}^n) \geqslant (1.173\ldots + o(1))^n, \\
\overline{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); k) &\leqslant \overline{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); k) \leqslant (3 + o(1))^{kn},
\end{aligned}$$

and that's all. In Sect. 7, we shall speak about very subtle "conditional" improvements to these bounds.

As for the two other quantities, they are subject to a general estimate (see [82]):

$$(c_1 k)^{c_2 n} \leqslant \overline{\chi}_{\mathbb{R}}((\mathbb{Q}^n, l_2); k) \leqslant \overline{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); k) \leqslant (3 + o(1))^{kn}.$$

Moreover, for fixed values of k, there was a series of consecutive publications (by this author, by I.M. Shitova (Mitricheva), by V.Y. Protassov, and so forth) with nontrivial lower bounds (see, e.g., [45, 105]). All such bounds have the form $(c+o(1))^n$ and we know that the corresponding upper bounds have the same form (with another constant). So it is natural to introduce the values $\tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); k)$, $\tilde{\chi}_{\mathbb{R}}((\mathbb{Q}^n, l_2); k)$, which are defined as the suprema of constants c in any lower bound. The current "records" are found in [45]. They are based on the consideration of distance graphs whose vertices are vectors from $\mathbb{Z}^n$ with prescribed numbers of coordinates of a given value:

$$V = \{\, \mathbf{x}:\ x_i \in \{0, 1, 2, 3, \ldots, d\},\ |\{i:\ x_i = j\}| = l_j,\ l_0 + \ldots + l_d = n\}.$$

Below we list some of them:

k	1	2	3	4	5	6	7	8	9	10	11	12
$\tilde{\chi} \geqslant$	1.239	1.465	1.667	1.848	2.013	2.165	2.308	2.442	2.570	2.691	2.807	2.919

The same type of work was also done for the l_1-metric (see [99, 105]).

Finally, the measurable chromatic numbers with many forbidden distances were studied. In [38], the best results concerning this subject are given.

3.4.2 Infinite Sequences of Forbidden Distances

Let's consider only the Euclidean spaces $\mathbb{R}^n$. However, let's take an infinite *monotone increasing sequence* of forbidden distances $\mathcal{A} = \{a_1, a_2, \dots\}$ and assume that $\frac{a_{i+1}}{a_i} \geqslant 1 + \frac{1}{d}$; i.e., the growth of the sequence is at least exponential. Such a sequence is called *lacunary*, and d is called the *lacunarity coefficient*. Denote by $\chi(n,d)$ the value

$$\chi(n,d) = \max_{\mathcal{A}} \chi((\mathbb{R}^n, l_2); \mathcal{A}),$$

where the maximum is taken over all possible lacunary sequences with lacunarity coefficient d.

Let $n = 1$. Even in this case, everything is quite nontrivial. The problem of finding $\chi(1,d)$ was posed in 1987 by Erdős (see [58]). Moreover, Erdős was just wondering whether $\chi(1,d)$ is finite. A positive answer to the question was given by Y. Katznelson in the early 1990s, but his paper appeared only in 2001 (see [58]). In fact, Katznelson proved the bound $\chi(1,d) \leqslant cd^2 \ln d$ with some $c > 0$.

During the last 10 years, several consecutive improvements of Katznelson's estimate were published. There was a paper by I.Z. Rusza, Z. Tuza, and M. Voigt; a paper by R. Akhunzhanv and N.G. Moshchevitin; and papers by A. Dubickas. But the current "record" is due to Y. Peres and W. Schlag (see [80]): $\chi(1,d) = O(d \ln d)$.

Surprisingly, no nontrivial lower estimates for $\chi(1,d)$ are known. Of course, $\chi(1,d) \geqslant d+2$, since we can consider the complete distance graph on the vertices $0, 1, \dots, d+1$ with any lacunary sequence of forbidden distances starting from $1, 2, \dots, d+1$. Using some elementary tools, one can replace $d+2$ by something like $1.5d$, and that's all what we know.

The problem of finding the value $\chi(1,d)$ is very close to some deep problems in Diophantine approximation (see [80]).

As for $n > 1$, *nothing* is known. The measurable variant $\chi_m(n,d)$ is infinite, but what about the initial one? If $\chi(2,d)$ is finite, then the difference between $\chi_m(2,d)$ and $\chi(2,d)$ is even more spectacular than the difference between $\chi(\mathbb{R}^2)$ and $\chi_m(\mathbb{R}^2)$.

3.4.3 Intervals of Forbidden Distances

We have already considered the problem of finding $\chi(\mathbb{R}^n;[1,d])$ in Sects. 2.1 and 2.2. The same problem was thoroughly studied by Kupavskii in [63] and [64] for $n \to \infty$. His results are as follows. Let K be an arbitrary convex centrally symmetric body in $\mathbb{R}^n$. Let ρ be the corresponding norm.

1. $\chi((\mathbb{R}^n,\rho);[1,d]) \leqslant (2(d+1)+o(1))^n$.
2. If $q > 2$, then $\chi((\mathbb{R}^n,l_q);[1,d]) \leqslant (2^{c_q}(d+1)+o(1))^n$. Here c_q is the same as in item 2 from Sect. 3.3.
3. If $d \geqslant 2$, then $\chi((\mathbb{R}^n,\rho);[1,d]) \geqslant \left(\frac{d}{2}\right)^n$.
4. If $d \geqslant 2$, then $\chi((\mathbb{R}^n,l_q);[1,d]) \geqslant (b\cdot d)^n$, where $b = \frac{\sqrt[p]{2}}{2}$ and $p = \max\left\{q, \frac{q}{q-1}\right\}$.
5. If $d \geqslant 2$, then $\chi((\mathbb{R}^n,l_2);[1,d]) \geqslant (b\cdot d)^n$, where $b \approx 0.755\cdot\sqrt{2}$.

Note that item 2 entails item 2 from Sect. 3.3 when $d = 1$ and the interval $[1,d]$ degenerates into one point $\{1\}$. Also, item 1 is a rough form of item 1 from Sect. 3.3.

3.5 The Chromatic Numbers of Spheres

Another important and deeply studied metric space is (S_r^{n-1}, l_2), where S_r^{n-1} is an $(n-1)$-dimensional sphere in $\mathbb{R}^n$ having radius r. Here we consider the chromatic number

$$\chi(S_r^{n-1}) = \chi((S_r^{n-1}, l_2);\{1\}).$$

Obviously, r is assumed to be greater than or equal to $1/2$, since otherwise $\chi(S_r^{n-1}) = 1$. Moreover, $\chi(S_{1/2}^{n-1}) = 2$ (see Sect. 1 for a similar comment).

In 1981, Erdős conjectured that for any $r > 1/2$, $\chi(S_r^{n-1}) \to \infty$ as $n \to \infty$ (see [31]). This conjecture was quickly proved by Lovász (see [72]), who used some topological methods (cf. [76]) and showed that for any $n \in \mathbb{N}$ and $r > 1/2$, $\chi(S_r^{n-1}) \geqslant n$. In the same paper, Lovász also makes the following assertion: *If $r < \sqrt{\frac{n}{2n+2}} \sim \frac{1}{\sqrt{2}}$, i.e., the length of any side of a regular n-simplex inscribed into S_r^{n-1} is smaller than 1, then* $\chi(S_r^{n-1}) \leqslant n+1$. This assertion is widely quoted (see, e.g., [115]). However, it is completely wrong. Lovász's idea (and his mistake) is, in some sense, natural. The way of reasoning could be as follows. Take the sphere S_r^{n-1} with $r < \sqrt{\frac{n}{2n+2}}$. Inscribe a regular n-simplex into this sphere. Let $\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}$ be its vertices. Let O be the center of S_r^{n-1}. Consider $n+1$ equal multidimensional angles generated by O and some n points from the set $\{\mathbf{x}_1,\ldots,\mathbf{x}_{n+1}\}$. Color the intersection of the sphere with the ith angle in the ith color. Since the side length of the simplex, *which is the diameter of any monochromatic part of the sphere*, is less than 1, we have a coloring avoiding distance 1. The mistake is just in italic text. Of course, the diameters of monochromatic parts are *not* attained on the sides of the simplex. For example, if $n = 2k$, then the maximum distance is between appropriately normed $\mathbf{x}_1 + \cdots + \mathbf{x}_k$ and $\mathbf{x}_{k+1} + \cdots + \mathbf{x}_{2k}$.

Recently, this author used the linear algebra method and succeeded in proving the following result (see [91] and [92]): *For any $r > \frac{1}{2}$, there exist a constant $\gamma = \gamma(r) > 1$ and a function $\varphi(n) = \varphi(n,r) = o(1)$ as $n \to \infty$, such that for every $n \in \mathbb{N}$, the inequality*

$$\chi(S_r^{n-1}) \geqslant (\gamma + \varphi(n))^n$$

holds. This result means, in particular, that for any $r > 1/2$, there exists an n_0 such that for $n \geqslant n_0$, $\chi(S_r^{n-1}) > n+1$. Moreover, the real growth of the chromatic number is always exponential, not linear.

Fixed values of r may even be replaced by sequences r_n of radii. Since the primality is naturally included in the linear algebra method (see Sect. 2.5), the corresponding results also depend on some facts/conjectures concerning the distribution of primes among naturals. For example, it is known that there exists a function $f(x) = O\left(x^{0.525}\right)$ such that for every $x > 0$, there is a prime number between x and $x + f(x)$ (see [10]). And it is conjectured that one can also take $f(x) = O\left(\ln^2 x\right)$. If we do not believe in this conjecture, then we certainly can show that (see [92]) *there exist constants c, n_0 such that for any sequence of radii satisfying the inequality*

$$r_n \geqslant \frac{1}{2} + \frac{c}{n^{0.475}}$$

and for any $n \geqslant n_0$, $\chi(S_{r_n}^{n-1}) > n+1$. If the conjecture is true, we can write $\sqrt{\frac{\ln n}{n}}$ instead of $\frac{1}{n^{0.475}}$ (see [92]).

Only for $r_n = \frac{1}{2} + O\left(\frac{1}{n}\right)$ does the estimate $\chi(S_{r_n}^{n-1}) \leqslant n+1$ hold. This is done just by inscribing a regular simplex into the sphere and carefully calculating the diameters of the corresponding monochromatic parts.

Some other upper estimates apply to $\chi(S_r^n)$. First of all, it is clear that for any r_n,

$$\chi(S_{r_n}^n) \leqslant \chi(\mathbb{R}^n) \leqslant (3 + o(1))^n. \tag{6}$$

Further, Rogers proved in [108] that any sphere of radius r in $\mathbb{R}^n$ can be covered by $\left(\frac{r}{\rho} + o(1)\right)^n$ spheres of radius $\rho < r$. In our case, this means that

$$\chi(S_{r_n}^{n-1}) \leqslant (2r_n + o(1))^n.$$

If $r_n < 3/2$, then this bound is better than that in (6).

More precisely, Rogers' estimate is as follows: *There is an absolute constant $c > 0$ such that, if $r > \frac{1}{2}$ and $n \geqslant 9$, any n-dimensional spheres of radius r can be covered by less than $cn^{5/2}(2r)^n$ spheres of radius $\frac{1}{2}$.* Such a precise formulation is not useful when r is a constant, but returning to $r_n \to \frac{1}{2}$, we may carefully apply this statement in order to obtain upper bounds like

$$\chi(S_{r_n}^{n-1}) \leqslant 2cn^{5/2}(2r_n)^n = \Theta\left(n^{5/2}(2r_n)^n\right). \tag{7}$$

Here the factor 2 is due to the fact that $\chi(S_{1/2}^{n-1}) = 2$. One should not forget that if, for example, $r_n = \frac{1}{2} + \Theta\left(\frac{1}{n}\right)$, then $(2r_n)^n = \Theta(1)$, so that estimate (7) is very good.

It is possible to evaluate even more sophisticated bounds for $\chi(S_{r_n}^{n-1})$, but this is not so interesting.

In small dimensions, the quantity $\chi(S_r^n)$ is discussed in [61, 112, 113], and here we do not dwell on this subject.

4 High Girth/Small Clique Number and High Chromatic Number

In Sect. 2.1, we already discussed a question on the existence of distance graphs with a high girth and chromatic number. Analogous questions in three dimensions were studied by O.I. Rubanov in [110], but much more interesting are the corresponding statements when $n \to \infty$.

Let's denote by $\omega(G)$ *the clique number* of a graph G, i.e., the maximum number of pairwise adjacent vertices in G (the maximum cardinality of a *clique* in G).

The results from Sect. 3.5 show implicitly that there exist distance graphs with a high chromatic number (growing exponentially in n) and not containing, say, triangles. Indeed, if a graph contains a triangle, then the minimum radius of a sphere, on which this graph could lie, is $\frac{1}{\sqrt{3}}$ (which is the radius of the circumference around a regular triangle with side length 1). But we know that $\chi(S_r^{n-1}) > c^n$ even for $r < \frac{1}{\sqrt{3}}$, and this is equivalent to the existence of a distance graph $G \subset S_r^{n-1}$ with $\chi(G) > c^n$ and $\omega(G) < 3$.

Thus, one can see that for any $k > 2$, there exist a constant $c_k > 1$ and a function $\delta(n) = o(1)$, $n \to \infty$, such that for every $n \in \mathbb{N}$, there is a distance graph $G \subset \mathbb{R}^n$ with $\omega(G) < k$ and $\chi(G) \geqslant (c_k + \delta(n))^n$. It is natural to try to optimize over the values of c_k for given k's. One could expect that the optimal constants would tend to $1.239\ldots$ as $k \to \infty$. So define

$$\zeta_{\text{clique}}(k) = \sup\{\zeta : \exists\, \delta(n) = o(1), \forall\, n, \exists\, G \text{ in } \mathbb{R}^n,\ \omega(G) < k,\ \chi(G) \geqslant (\zeta + \delta(n))^n\}.$$

The quantity $\zeta_{\text{clique}}(k)$ was introduced in [93], and since then it has been intensively studied (see [24, 103, 104]). Below we present some best-known estimates for $\zeta_{\text{clique}}(k)$.

k	$\leqslant 5$	6	7	8	9	10	11	12
$\zeta_{\text{clique}}(k) \geqslant$	1.058	1.074	1.085	1.093	1.099	1.103	1.107	1.109
k	13	14	15	16	17	100	1000	10^6
$\zeta_{\text{clique}}(k) \geqslant$	1.112	1.115	1.122	1.128	1.133	1.219	1.237	1.239

The tabular shows that, indeed, $\zeta_{\text{clique}}(k) \to 1.239\ldots$ as $k \to \infty$. The ideas of proving such estimates are in a combination of the linear algebra, some coding theory (see [74]), and probability (see [8]).

A quantity $\zeta_{\text{clique}}(k,m)$ was also studied in [24], where m means the number of forbidden distances. We do not dwell on this here.

Now, what about high girth, not clique number? It is not difficult to prove that any of the graphs, to which we can apply the linear algebra, does contain cycles of arbitrary even length. So it remains to consider the values

$$\zeta_{\text{odd girth}}(k) = \sup\{\zeta : \exists\, \delta(n) = o(1),\ \forall\, n,\ \exists\, G \text{ in } \mathbb{R}^n,\ g_{\text{odd}}(G) > k,\ \chi(G) \geqslant (\zeta + \delta(n))^n\},$$

where $g_{\text{odd}}(G)$ is the length of the shortest odd cycle in G. The estimate is proved? in [24]:

$$\zeta_{\text{odd girth}}(2k+1) \geqslant 2 \cdot \left(\frac{k}{2k+1}\right)^{\frac{k}{2k+1}} \cdot \left(\frac{k+1}{2k+1}\right)^{\frac{k+1}{2k+1}}.$$

Recently, this author and Kupavskii found many important improvements to results from this section, but our paper [1, 2].

5 Random Distance Graphs Related to the Chromatic Number

The classical Erdős–Rényi theory of random graphs consists of studying the probability spaces $G(n,p) = (\Omega_n, \mathcal{F}_n, P_{n,p})$, where Ω_n is the set of all possible graphs on the set $V_n = \{1,\ldots,n\}$ of vertices ($|\Omega_n| = 2^{\binom{n}{2}}$), $\mathcal{F}_n = 2^{\Omega_n}$, and for $G = (V_n, E)$,

$$P_{n,p}(G) = p^{|E|}(1-p)^{\binom{n}{2}-|E|}.$$

In other words, we take the complete graph K_n and delete its edges independently, each with probability $1-p$. The theory started in the 1950s, and now it has been deeply elaborated (see [8, 12, 33–35]).

In the previous section, we wrote that one should combine some linear algebra and probability in order to obtain some estimates for the chromatic numbers of distance graphs. A general idea is as follows. Take a complete *distance* graph with some good coloring properties (say, having a large chromatic number) and then start deleting its edges randomly. Prove eventually that with high probability, the chromatic number is still large, but cliques (cycles) have disappeared.

A natural example of a complete distance graph with "good coloring properties" is suggested by the Frankl–Wilson theorem. Let $n = 4k$ and $G = (V,E)$, where

$$V = \{\, \mathbf{x} = (x_1,\ldots,x_n) : x_i \in \{0,1\},\ x_1 + \ldots + x_n = 2k\},\quad E = \{\{\,\mathbf{x},\ \mathbf{y}\} : (\,\mathbf{x},\ \mathbf{y}) = k\}.$$

Denote the quantity $|V|$ by N. Define a probability space $G_{\rm dist}(N,p)$ just in the same way as the space $G(n,p)$ was defined. In other words, take any spanning subgraph $H=(V,F)$ of the complete distance graph with probability $p^{|F|}(1-p)^{|E|-|F|}$.

During the last several years, some asymptotic properties of the random graph $G_{\rm dist}(N,p)$ have been studied. On the one hand, connected components of this graph were investigated. A.R. Yarmukhametov found a series of such results in [119]. In particular, he found threshold functions for connectivity of random distance graphs and for the property of containing a "giant" component.

On the other hand, some (0,1)-laws for the random distance graph were established. It is known that for the Erdős–Rényi model, the following is true: *If $p=p(n)$ is such that for every $\alpha>0$, $\min\{p,1-p\}\cdot n^{\alpha}\to\infty$ as $n\to\infty$, and $\mathcal{A}$ is any graph property that can be expressed in the first-order language, then* $\lim\limits_{n\to\infty} P_{n,p}(\mathcal{A})\in\{0,1\}$. This result was proved in [36] and [44] (see also [8]). Recently, M.E. Zhukovskii showed that the random distance graph is not subject to a (0,1)-law. At the same time, he found some "weak" (0,1)-laws, which are true for the distance graphs, and he proved the classical law for a subsequence of random distance graphs (see [120–122]).

6 The Independence Numbers of Some Distance Graphs

6.1 A Possible Way to Improve Lower Bounds for the Chromatic Numbers

As we know, all the best lower estimates for the chromatic numbers of various spaces were obtained with the help of distance graphs whose sets of vertices belonged to $\mathbb{Z}^n$. Moreover, in the case of one forbidden distance, only (0,1)- and (-1,0,1)-vectors gave strong results (cf. Sect. 2.5). In this section, we shall discuss this problem in depth.

In order to get lower bounds for the chromatic numbers, we always use the inequality $\chi(G)\geqslant\frac{|V|}{\alpha(G)}$. This is quite natural and, in some sense, optimal. The point is that "almost all" graphs have the property $\chi(G)\sim\frac{|V|}{\alpha(G)}$ (see [8, 12]). So the main problem is in finding tight upper estimates for the independence numbers $\alpha(G)$.

If considering (0,1)-graphs like in (5), then one can easily see that the linear algebra upper bounds are asymptotically sharp in order. Explicit constructions providing almost the same lower bounds are quite simple. For example, if $a=3$, $b=2$ (see Sect. 2.5), then one can take all the vectors with $x_1=x_2=1$. The number of such vectors is $n-2$, which is very close to the upper estimate $\alpha(G_{3,2})\leqslant n$ proved in Sect. 2.5. For $a=5$, $b=\sqrt{6}$, one should fix the three first coordinates. Then $\binom{n-2}{2}$ is really not far from $\binom{n}{2}+\binom{n}{1}+1$. In the general case, one may apply the famous Erdős–Ko–Rado theorem (see [32]) and its extensions by Frankl, Wilson, Ahlswede, and Khachatrian (see [4–6]). In any case, upper and lower estimates do not differ substantially.

However, in the case of $(-1,0,1)$-graphs, everything is much worse. Let's just consider an example. Set $n = 4k$ and $G = (V,E)$, with

$$V = \{\, \mathbf{x} = (x_1,\ldots,x_n) : \, x_i \in \{-1,0,1\}, |\{i : \, x_i = 1\}| = |\{i : \, x_i = -1\}| = k\},$$
$$E = \{\{\, \mathbf{x}, \, \mathbf{y}\} : \, (\, \mathbf{x}, \, \mathbf{y}) = 0\}.$$

Here linear algebra gives the bound $\alpha(G) \leqslant (2.462\ldots + o(1))^n$. Nevertheless, nontrivial constructions [generalizing, in some sense, the Erdős–Ko–Rado theorems to $(-1,0,1)$-vectors] only provide the estimate $\alpha(G) \geqslant (2.264\ldots + o(1))^n$. In the papers [47, 73, 78], many such constructions are proposed. It is also conjectured there that the lower estimates are tight and the upper ones are not. In other words, the linear algebra method probably begins to fail when, from the (0,1)-situation, we proceed to the (-1,0,1)-case. Some computer-based arguments also support this conjecture.

If the just-mentioned conjecture is true, then one can again optimize over different choices of parameters (numbers of 0s, 1s, -1s, etc.), substantially improving all the lower estimates for various chromatic numbers from the previous sections. Unfortunately, no techniques have been developed that are better than the linear algebraic one.

6.2 Independent Sets and Measurable Chromatic Numbers

Another well-studied subject linking independent sets and colorings concerns the measurable chromatic numbers. Consider a Lebesgue-measurable set $A \subset \mathbb{R}^n$. Define its *upper density* as follows:

$$\overline{\delta}(A) = \overline{\lim_{r\to\infty}} \frac{\mu\,(A \cap B_n(r))}{\mu(B_n(r))},$$

where μ is the Lebesgue measure and $B_n(r)$ is the n-dimensional ball of radius r centered at the origin. Denote by $m_1(n)$ the quantity

$$m_1(n) = \sup\left\{\overline{\delta}(A) : A \subset \mathbb{R}^n, A \text{ avoids unit distance}\right\}.$$

This quantity is a natural analogue of the independence number, and, of course, $\chi_m(\mathbb{R}^n) \geqslant \frac{1}{m_1(n)}$.

The following table shows the best-known upper and lower bounds for $m_1(n)$, $2 \leqslant n \leqslant 8$. Note that for $n \geqslant 3$, all the lower estimates for $\chi_m(\mathbb{R}^n)$ cited in Sect. 2.2 are obtained using the results from this table.

n	Upper bound for $m_1(n)$	Lower bound for $m_1(n)$
2	0.26841	0.2293
3	0.16560	0.09877
4	0.11293	0.04413
5	0.07528	0.01833
6	0.05157	0.00806
7	0.03612	0.00352
8	0.02579	0.00165

All the upper bounds are obtained in [38]. The lower bound for $n = 2$ is due to Croft (see [23]), and the other lower bounds were recently found by this author, Kupavskii, and Titova (see [102]).

7 Conditional Results on the Chromatic Numbers

From Sect. 23.6.1, we know that probably the linear algebra method is not the best possible method. Inspired by this knowledge, this author proposed in 2003 an idea that he called the "alteration principle" (see [94]). Roughly speaking, the idea is as follows. Take a distance graph $G_{a_1,a_2} = (V, E_{a_1,a_2})$ (say, with two forbidden distances a_1, a_2), whose independence number can be bounded from above by some s with the help of linear algebra. Forget about a_2. For the new graph $G_{a_1} = (V, E_{a_1})$ (with one forbidden distance a_1), linear algebra does not work. However, one may *assume* that $\alpha(G_{a_1}) < t$ with some t. If the assumption is true, then $\chi(G_{a_1}) > \frac{|V|}{t}$. Otherwise, take an independent set $W \subset V$ such that $|W| \geqslant t$. Consider the graph $G_{a_2} = (W, E_{a_2})$ with one forbidden distance a_2. Then $\chi(G_{a_2}) \geqslant \frac{t}{s}$. Clearly, $\chi(\mathbb{R}^n) \geqslant \max\{\chi(G_{a_1}), \chi(G_{a_2})\}$. Thus, for any t, *either* $\chi(\mathbb{R}^n) \geqslant \frac{|V|}{t}$ *or* $\chi(\mathbb{R}^n) \geqslant \frac{t}{s}$ (here we consider two *alternatives*). Thus, anyway,

$$\chi(\mathbb{R}^n) \geqslant \max_t \min\left\{\frac{|V|}{t}, \frac{t}{s}\right\}.$$

This idea does not work exactly for $\chi(\mathbb{R}^n)$, but in some other situations, it helps obtain "conditional" results. In Sect. 3.4.1, we discussed the values

$$\tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); k), \quad \tilde{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); k), \quad \tilde{\chi}_{\mathbb{R}}((\mathbb{Q}^n, l_2); k), \quad \tilde{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); k).$$

In [105], the following three estimates were proved:

$$\limsup_{n\to\infty} \tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 2) + \limsup_{n\to\infty} \tilde{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); 2) \geqslant 1.465\ldots + 1.173\ldots + 0.043\ldots,$$

$$\limsup_{n\to\infty} \tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 1) + \limsup_{n\to\infty} \tilde{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); 1) \geqslant 1.239\ldots + 1.173\ldots + 0.005\ldots,$$

$$\limsup_{n\to\infty} \tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 2) + \limsup_{n\to\infty} \tilde{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); 2) \geqslant 1.465\ldots + 1.239\ldots + 0.008\ldots.$$

The results are very funny and subtle. Without "alterations," we can only show that

$$\tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 2) + \tilde{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); 2) \geqslant 1.465\ldots + 1.173\ldots,$$
$$\tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 1) + \tilde{\chi}_{\mathbb{Q}}((\mathbb{Q}^n, l_2); 1) \geqslant 1.239\ldots + 1.173\ldots,$$
$$\tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 2) + \tilde{\chi}_{\mathbb{Q}}((\mathbb{R}^n, l_2); 2) \geqslant 1.465\ldots + 1.239\ldots.$$

The additional constants look strange and unexpected. Either one chromatic number or the other admits a much better lower bound. One should only take a subsequence of the dimensions.

There are other applications of alterations. Some of them can be found in [94]. Some of them concern lacunary sequences of forbidden distances (see [77]). And some of them will be discussed in Sect. 8.5.

8 Borsuk's Problem

8.1 "Small" Dimensions

For $n = 1$, Borsuk's conjecture is trivially true. Of course, $f(1) \geqslant 2$, and $f(1) \leqslant 2$, since every set of diameter 1 on the real line can be covered by a segment of length 1, which can be divided into two parts each having diameter $1/2$.

For $n = 2$, Borsuk himself proved in [14] that $f(2) \leqslant 3$ (the opposite inequality is just given by a regular triangle). This is done by covering an arbitrary set of diameter 1 in the plane by a regular hexagon with distance 1 between parallel sides. On the other hand, in 1946, Erdős found a purely combinatorial proof of the fact that any (finite) graph of diameters in the plane has a chromatic number not exceeding 3 (see [30]).

The first proof of Borsuk's conjecture in $\mathbb{R}^3$ belongs to Eggleston (see [28]). Other proofs, which improve Eggleston's result in some sense, continue to appear (see [85] for the details). The bound $\chi(G) \leqslant 4$, which is true for any finite three-dimensional graph of diameters, was also proved by Heppes and Révész in [50].

In dimension 4, only the bound $f(4) \leqslant 9$ due to Lassak (see [69]) is known. However, no counterexamples in this dimension have also been found.

8.2 Some General Results

First of all, there are some partial solutions to Borsuk's problem. We list them here.

1. If $\Omega \subset \mathbb{R}^n$ has C^1-boundary, then $f(\Omega) \leqslant n+1$ (see [49]).
2. If $\Omega \subset \mathbb{R}^n$ is centrally-symmetric, then $f(\Omega) \leqslant n+1$ (see [60]).

3. If $\Omega \subset \mathbb{R}^n$ is invariant under the action of the group of symmetries of a regular n-simplex, then $f(\Omega) \leqslant n+1$ (see [109]).
4. If $\Omega \subset \{0,1\}^n$, $n \leqslant 9$, then $f(\Omega) \leqslant n+1$ (see [123]).

Also, some upper bounds for $f(n)$ are known. The best ones of them are due to Lassak (see [69]), Schramm (see [111]), and Bourgain and Lindenstrauss (see [15]). Lassak's bound is by $2^{n-1}+1$. Schramm's and Bourgain–Lindenstrauss' bounds are by $\left(\sqrt{\frac{3}{2}}+o(1)\right)^n$. So the first bound is better in fixed dimensions [for example, $f(4) \leqslant 9$, cf. Sect. 8.1], but the second one beats it when $n \to \infty$.

8.3 Counterexamples to Borsuk's Conjecture

The first counterexample was constructed by Kahn and Kalai in [56]. Surprisingly, it is based on a complete distance graph, which is almost the same as the graph from Sect. 5. Roughly speaking, the idea is as follows. Let $n = 4p$, where p is prime. Take the set

$$V = \{\, \mathbf{x} = (x_1, \ldots, x_n) : \; x_i \in \{-1,1\}, |\{i : \; x_i = 1\}| = |\{i : \; x_i = -1\}| = 2p\}.$$

To each vector $\mathbf{x}$ from V, assign the vector $\mathbf{x}^*$ whose coordinates are products $x_i x_j$ $(i,j = 1,\ldots,n)$ of the coordinates of $\mathbf{x}$. It is easy to see that the diameter of the set $V^* \subset \mathbb{R}^{n^2}$, which consists of all the vectors $\mathbf{x}^*$, is attained on some $\mathbf{x}^*$, $\mathbf{y}^*$ if and only if $(\mathbf{x},\mathbf{y}) = 0$. So by the Frankl–Wilson theorem, $f(n^2) \geqslant (c+o(1))^n$, which means that $f(n) \geqslant (c'+o(1))^{\sqrt{n}}$.

In the work of Kahn and Kalai, $c' = 1.203\ldots$ and the first dimension, where Borsuk's conjecture fails, is $n = 2015$. Now we know that (see [95])

$$f(n) \geqslant \left(\left(\frac{2}{\sqrt{3}}\right)^{\sqrt{2}} + o(1)\right)^{\sqrt{n}} = (1.2255\ldots + o(1))^{\sqrt{n}}$$

and that $f(n) > n+1$ for $n \geqslant 298$ (see [107]).

Thus, the gap between upper and lower estimates for $f(n)$ is still huge, and nothing is known about $n \in \{4,\ldots,297\}$. Item 4 from Sect. 8.2 shows that one should not expect (0,1)-counterexamples for $n \leqslant 9$, but that's all. We conjecture that $f(4) = 5$ and $f(n) > n+1$ for $n \geqslant 5$. However, this conjecture is very far from its proof.

8.4 Counterexamples on Spheres of Small Radii

The diameter of the set V^* from Sect. 8.3 equals $\sqrt{2n^2}$. At the same time, one can easily see that the radius of the smallest sphere S such that $V^* \subset S$ is equal to n.

So if we normalize the construction to get a graph of diameters with diameter 1, then the radius becomes equal to $\frac{1}{\sqrt{2}}$.

We know (see [55] and cf. Sect. 3.5) that *any* set of diameter 1 can be covered by a sphere of radius $\frac{1}{\sqrt{2}}$. So the counterexample from Sect. 8.3 is the "fattest" possible. In fact, all the counterexamples, which were obtained until 2010, had the same "girth." This is quite natural, since, as usual, it is much easier to get a graph with a high chromatic number by including large cliques in the graph. In our case, cliques are regular simplices (preferably, of maximum dimension), so that to cover our graphs of diameters by spheres means to cover at least one simplex by a sphere, which can be done only when the radius is not less than $\sqrt{\frac{n}{2n+2}} \sim \frac{1}{\sqrt{2}}$.

However, Sect. 3.5 showed us that sometimes any intuition can be broken. Thus, one can probably find counterexamples to Borsuk's conjecture, which are finite graphs of diameters lying on spheres whose radii are smaller than $\frac{1}{\sqrt{2}}$? Of course, such results would be even more surprising and complicated than the ones concerning distance graphs, but why not?

In the paper [96], this author introduced the value $f_r(n) = \max f(\Omega)$, where the maximum is taken over all possible sets $\Omega \subset S_r^{n-1}$ such that $\mathrm{diam}\Omega = 1$. Here, of course, $r \geqslant \frac{1}{2}$. By the Borsuk–Ulam theorem (cf. Sect. 1), $f_{1/2}(n) = n+1$. By the above-mentioned results, $f_{1/\sqrt{2}}(n) \geqslant (1.2255\ldots + o(1))^{\sqrt{n}}$. Now, the question is whether or not $f_r(n) > n+1$, provided $r < \frac{1}{\sqrt{2}}$. And in [96], this author proved the following theorem: *For any* $r > \sqrt{\frac{3}{8}} = 0.612\ldots$*, there exists an* n_0 *such that for every* $n \geqslant n_0$*,* $f_r(n) > n+1$.

The method in [96] was very close to the one described in Sect. 8.3. However, it was exhausted completely, since for $r \approx \sqrt{\frac{3}{8}}$, it gave estimates only by some "epsilon" greater than $n+1$. For $r = \sqrt{\frac{3}{8}}$, it failed.

Recently, this author, together with Kupavskii, made another attempt to solve the problem of bounding $f_r(n)$ from below, and the joint efforts were much more successful. Here are some of the obtained results (see [66]).

1. For any $r > \frac{1}{2}$, there exist numbers $k = k(r) \in \mathbb{N}$, $c = c(r) > 1$ and a function $\delta(n) = o(1)$ such that
$$f_r(n) \geqslant (c + \delta(n))^{\sqrt[2k]{n}}.$$
2. Let $r_n = \frac{1}{2} + \varphi(n)$, where $\varphi = o(1)$ and $\varphi(n) \geqslant c\frac{\ln\ln n}{\ln n}$ for all n and a large enough $c > 0$. Then there exists an n_0 such that for $n \geqslant n_0$, $f_{r_n}(n) > n+1$.
3. Let $r_n = \frac{1}{2} + \varphi(n)$, where $\varphi = O(1/n)$. Then $f_{r_n}(n) \leqslant n+1$.

Item 1 means that counterexamples to Borsuk's conjecture can lie on spheres of any radius $r > \frac{1}{2}$. Moreover, the type of a lower bound for $f_r(n)$ is similar to the one for $f(n)$. Only, instead of $\sqrt{n}$, we have $\sqrt[2k]{n}$ with a constant k depending on the proximity of r to $1/2$.

Item 2 says that a constant r may even be replaced by a sequence r_n whose elements differ from $1/2$ by an $o(1)$. Notice that in Sect. 3.5, we had almost the same situation for distance graphs on spheres. However, in that case the value of $o(1)$ was something like $\frac{1}{n^{\alpha}}$, and here we have $o(1) = \frac{\ln\ln n}{\ln n}$. This is due to many difficulties coming from the more complicated structure of graphs of diameters.

Item 3 can be proved by inscribing a regular simplex into $S_{r_n}^{n-1}$ (see Sect. 3.5).

8.5 Alterations

In Sect. 7, we discussed an alteration principle introduced by this author in [94], and we talked about its applications to various chromatic numbers. In [97] this author noticed that the same principle can be applied in an even more unusual way. Let $\tilde{f} = \tilde{f}(n)$ be the greatest function in a bound $f(n) \geqslant (\tilde{f} + o(1))^{\sqrt{n}}$. We certainly know that $\tilde{f} \geqslant 1.2255\ldots$, but, in principle, $\tilde{f}$ may grow (cf. Sect. 8.2). Let $\tilde{\chi} = \tilde{\chi}_{\mathbb{R}}((\mathbb{R}^n, l_2); 1)$ (see Sects. 3.4.1 and 7). Then, of course, $\tilde{\chi} \geqslant 1.239\ldots$. This author proved in [98] that

$$\limsup_{n\to\infty} \tilde{f} + \limsup_{n\to\infty} \tilde{\chi} \geqslant 1.2255\ldots + 1.239\ldots + 0.004\ldots.$$

This result was improved in [106]:

$$\limsup_{n\to\infty} \tilde{f} + \limsup_{n\to\infty} \tilde{\chi} \geqslant 1.2255\ldots + 1.239\ldots + 0.017\ldots.$$

So either the known lower estimate for $\chi(\mathbb{R}^n)$ can be substantially improved or the same can be done for $f(n)$.

8.6 Coming Back to Small Dimensions

Let Φ be an arbitrary bounded set in $\mathbb{R}^m$. Take an $n \in \mathbb{N}$. Denote by $d_n^m(\Phi)$ the quantity

$$d_n^m(\Phi) = \inf\{x \in \mathbb{R}_+ : \ \Phi \subseteq \Phi_1 \cup \ldots \cup \Phi_n, \forall i \ \operatorname{diam} \Phi_i \leqslant x\}.$$

In other words, we want to decompose Φ into a fixed number of subsets having diameters that are as small as possible.

Let $d_n^m = \sup\limits_{\Phi} d_n^m(\Phi)$, where the supremum is taken over all sets Φ of diameter 1. The problem of finding the values d_n^m was initiated by Lenz in 1956 (see [70], [71]) and is apparently very close to Borsuk's problem.

The two-dimensional case is the most studied. Many results (old and new) can be found in a recent paper by V.P. Filimonov (see [39]). We'll just mention some of them:

n	1	2	3	4	5	6	7	8	9	10	11	12
$d_n^2 \geqslant$	1	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	0.5877	0.5051	0.5	0.4338	0.3826	0.342	$\frac{1}{3}$	$\frac{1}{3}$
$d_n^2 \leqslant$	1	1	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{2}}$	0.602	0.5577	0.5	0.4456	0.4047	0.4012	0.397	0.366

The three-dimensional case has also been investigated (see, e.g., [85, 101]). We do not dwell on this here.

For higher dimensions, the best results can be found in [100].

Finally, Filimonov succeeded in proving recently that for any Jordan measurable set Φ, there exists a limit of the sequence $\sqrt[m]{n} \cdot d_n^m(\Phi)$ (see [40, 41]).

Acknowledgements This work is done under the financial support of the following grants: Grant 12-01-00683 of Russian Foundation for Basic Research, Grant MD-8390.2010.1 of the Russian President, Grant NSh-2519.2012.1 supporting leading scientific schools in Russia.

References

1. A.B. Kupavskii, A.M. Raigorodskii, Distance Graphs with Large Chromatic Numbers and Small Clique Numbers, Doklady Mathematics, **85** No. 3, 394-398 (2012)
2. A.B. Kupavskii, Distance graphs with large chromatic number and arbitrary girth, Moscow J. of Comb. and Number Theory, 2(N2), 52–62 (2012)
3. P.K. Agarwal, J. Pach, *Combinatorial Geometry* (Wiley, New York, 1995)
4. R. Ahlswede, V.M. Blinovsky, *Lectures on Advances in Combinatorics* (Springer, Berlin, 2008)
5. R. Ahlswede, L.H. Khachatrian, The complete nontrivial-intersection theorem for systems of finite sets. J. Combin. Theor. A **76**, 121–138 (1996)
6. R. Ahlswede, L.H. Khachatrian, The complete intersection theorem for systems of finite sets. Eur. J. Combin. **18**, 125–136 (1997)
7. N. Alon, L. Babai, H. Suzuki, Multilinear polynomials and Frankl–Ray-Chaudhuri–Wilson type intersection theorems. J. Combin. Theor. A, **58**, 165–180 (1991)
8. N. Alon, J. Spencer, *The Probabilistic Method*, 2nd edn. (Wiley–Interscience Series in Discrete Mathematics and Optimization, New York, 2000)
9. L. Babai, P. Frankl, in *Linear Algebra Methods in Combinatorics*, Part 1, Department of Computer Science, University of Chicago, Preliminary version 2, September 1992
10. R.C. Baker, G. Harman, J. Pintz, The difference between consecutive primes, II. Proc. Lond. Math. Soc. **83**, 532–562 (2001)
11. M. Benda, M. Perles, Colorings of metric spaces. Geombinatorics **9**, 113–126 (2000)
12. B. Bollobás, *Random Graphs*, 2nd edn. (Cambridge University Press, Cambridge, 2001)
13. V.G. Boltyanski, H. Martini, P.S. Soltan, *Excursions into Combinatorial Geometry*, Universitext (Springer-Verlag, Berlin, 1997)
14. K. Borsuk, Drei Sätze über die *n*-dimensionale euklidische Sphäre. Fundamenta Math. **20**, 177–190 (1933)

15. J. Bourgain, J. Lindenstrauss, On covering a set in $\mathbb{R}^d$ by balls of the same diameter, ed. by J. Lindenstrauss, V. Milman. *Geometric Aspects of Functional Analysis*, Lecture Notes in Math., vol. 1469 (Springer-Verlag, Berlin, 1991), pp. 138–144
16. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, Berlin, 2005)
17. N.G. de Bruijn, P. Erdős, A colour problem for infinite graphs and a problem in the theory of relations. Proc. Koninkl. Nederl. Acad. Wet., Ser. A **54**(N5), 371–373 (1951)
18. K. Cantwell, Finite Euclidean Ramsey theory. J. Combin. Theor. A **73**(N2), 273–285 (1996)
19. K.B. Chilakamarri, The unit-distance graph problem: A brief survey and some new results. Bull. Inst. Comb. Appl. **8**, 39–60 (1993)
20. J. Cibulka, On the chromatic number of real and rational spaces. Geombinatorics **18**(N2), 53-66 (2008)
21. J.H. Conway, N.J.A. Sloane, *Sphere Packings, Lattices and Groups* (Springer-Verlag, Berlin, 1988)
22. D. Coulson, A 15-coloring of 3-space omitting distance one. Discrete Math. **256**, 83–90 (2002)
23. H.T. Croft, Incident incidents. Eureka (Cambridge) **30**, 22–26 (1967)
24. E.E. Demekhin, A.M. Raigorodskii, O.I. Rubanov, *Distance graphs having large chromatic numbers and not containing cliques or cycles of a given size*, to appear in **Sb**. Math.
25. P. O'Donnel, R. Hochberg, Some 4-chromatic unit-distance graphs without small cycles. Geombinatorics **5**(N4), 137–142 (1996)
26. P. O'Donnell, Arbitrary girth, 4-chromatic unit distnace graphs in the plane. Graph description. Geombinatorics **9**, 145–152 (2000)
27. P. O'Donnell, Arbitrary girth, 4-chromatic unit distance graphs in the plane. Graph embedding. Geombinatorics **9**, 180–193 (2000)
28. H.G. Eggleston, Covering a three-dimensional set with sets of smaller diameter. J. Lond. Math. Soc. **30**, 11–24 (1955)
29. P. Erdős, in *Unsolved Problems*. Congress Numerantium XV—Proceedings of the 5th British Comb. Conf. 1975, (1976), p. 681
30. P. Erdős, On sets of distances of n points. Am. Math. Monthly **53**, 248–250 (1946)
31. P. Erdős, R.L. Graham, in *Problem Proposed at the 6th Hungarian Combinatorial Conference*, Eger, Hungary, July 1981
32. P. Erdős, C. Ko, R. Rado, Intersection theorems for systems of finite sets. J. Math. Oxford, Sec. **12**(48) (1961)
33. P. Erdős, A. Rényi, On random graphs I. Publ. Math. Debrecen **6**, 290–297 (1959)
34. P. Erdős, A. Rényi, On the evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17–61 (1960)
35. P. Erdős, A. Rényi, On the evolution of random graphs. Bull. Inst. Int. Statist. Tokyo **38**, 343–347 (1961)
36. R. Fagin, Probabilities in finite models. J. Symbolic Logic **41**, 50–58 (1976)
37. K.J. Falconer, The realization of distances in measurable subsets covering $\mathbb{R}^n$. J. Combin. Theor. A **31**, 187–189 (1981)
38. F.M. de Oliveira Filho, F. Vallentin, Fourier analysis, linear programming, and densities of distance avoiding sets in $\mathbb{R}^n$. J. Eur. Math. Soc. **12**, 1417–1428 (2010)
39. V.P. Filimonov, Covering planar sets. Sb. Math. **201**(N8), 1217–1248 (2010)
40. V.P. Filimonov, Division codes and an area theorem. Dokl. Math. **79**(N3), 345–348 (2009)
41. V.P. Filimonov, Small deviation theorem for division codes. Dokl. Math. **81**(N2), 278–281 (2010)
42. P. Frankl, V. Rödl, Forbidden intersections. Trans. Am. Math. Soc. **300**(N1), 259–286 (1987)
43. P. Frankl, R. Wilson, Intersection theorems with geometric consequences . Combinatorica **1**, 357–368 (1981)
44. Yu.V. Glebskii, D.I. Kogan, M.I. Legon'kii, V.A. Talanov, Range and degree of realizability of formulas in the restricted predicate calculus. Cybernetics **5**, 142–154 (1969)
45. E.S. Gorskaya, I.M. Mitricheva, V.Y. Protasov, A.M. Raigorodskii, Estimating the chromatic numbers of Euclidean spaces by methods of convex minimization. Sb. Math. **200**(N6), 783–801 (2009)

46. P.M. Gruber, C.G. Lekkerkerker, *Geometry of Numbers* (North-Holland, Amsterdam, 1987)
47. A.E. Guterman, V.K. Lyubimov, A.M. Raigorodskii, A.S. Usachev, On the independence numbers of distance graphs with vertices at $\{-1,0,1\}^n$: estimates, conjectures, and applications to the Borsuk and Nelson–Erdős–Hadwiger problems. J. Math. Sci. **165**(N6), 689–709 (2010)
48. H. Hadwiger, Ein Überdeckungssatz für den Euklidischen Raum. Portugaliae Math. **4**, 140–144 (1944)
49. H. Hadwiger, Überdeckung einer Menge durch Mengen kleineren Durchmessers. Comm. Math. Helv. **18**, 73–75 (1945/46); Mitteilung betreffend meine Note: Überdeckung einer Menge durch Mengen kleineren Durchmessers. Comm. Math. Helv. **19**, 72–73 (1946/47)
50. A. Heppes, P. Révész, Zum Borsukschen Zerteilungsproblem. Acta Math. Acad. Sci. Hung. **7**, 159–162 (1956)
51. L.L. Ivanov, On the chromatic numbers of $\mathbb{R}^2$ and $\mathbb{R}^3$ with intervals of forbidden distances. Vestnik of the Russian University of People's Friendship (N1), 14–29 (2011) [in Russian]
52. L.L. Ivanov, An estimate of the chromatic number of the space $\mathbb{R}^4$. Russ. Math. Surv. **61**(N5), 984–986 (2006)
53. P.D. Johnson, Coloring abelian groups. Discrete Math. **40**, 219–223 (1982)
54. P.D. Johnson, Coloring the rational points to forbid the distance 1—a tentative history and compendium. Geombinatorics **16**, 209–218 (2006)
55. H.W.E. Jung, Über die kleinste Kugel, die eine räumliche Figur einschliesst. J. Reine und Angew. Math. **123**, 241–257 (1901)
56. J. Kahn, G. Kalai, A counterexample to Borsuk's conjecture. Bull. (new series) AMS **29**(N1), 60–62 (1993)
57. J.-H. Kang, Z. Füredi, Distance graphs on $\mathbb{Z}^n$ with l_1-norm. Theor. Comp. Sci. **319**(N1–3), 357–366 (2004)
58. Y. Katznelson, Chromatic numbers of Cayley graphs on Z and recurrence. Combinatorica **21**, 211–219 (2001)
59. V. Klee, S. Wagon, *Old and New Unsolved Problems in Plane Geometry and Number Theory* (Mathematical Association of America, Washington, DC, 1991)
60. R. Knast, An approximative theorem for Borsuk's conjecture. Proc. Cambridge Phil. Soc. (N1), 75–76 (1974)
61. A.B. Kupavskiy, On coloring spheres embedded into $\mathbb{R}^n$. Sb. Math. **202**(N6), 83–110 (2011)
62. A.B. Kupavskiy, On lifting of estimation of chromatic number of $\mathbb{R}^n$ in higher dimension. Dokl. Math. **429**(N3), 305–308 (2009)
63. A.B. Kupavskii, On the chromatic number of $\mathbb{R}^n$ with an arbitrary norm. Discrete Math. **311**, 437–440 (2011)
64. A.B. Kupavskii, The chromatic number of $\mathbb{R}^n$ with a set of forbidden distances. Dokl. Math. **82**(N3), 963–966 (2010)
65. A.B. Kupavskiy, A.M. Raigorodskii, On the chromatic number of $\mathbb{R}^9$. J. Math. Sci. **163**(N6), 720–731 (2009)
66. A.B. Kupavskii, A.M. Raigorodskii, Counterexamples to Borsuk's conjecture on spheres of small radii, submitted to Moscow J. Comb. and Number Theory. It is accepted and will be published in **2**(3) (2012). available at arXiv: 1010.0383
67. D.G. Larman, A note on the realization of distances within sets in Euclidean space. Comment. Math. Helvet. **53**, 529–535 (1978)
68. D.G. Larman, C.A. Rogers, The realization of distances within sets in Euclidean space. Mathematika **19**, 1–24 (1972)
69. M. Lassak, An estimate concerning Borsuk's partition problem. Bull. Acad. Polon. Sci. Ser. Math. **30**, 449–451 (1982)
70. H. Lenz, Zerlegung ebener Bereiche in konvexe Zellen von möglichst kleinern Durchmesser. Jahresbericht d. DMV Bd. **58**, 87–97 (1956)
71. H. Lenz, Über die Bedeckung ebener Punktmengen durch solche kleineren Durchmessers. Arch. Math. **VII**, 34–40 (1956)

72. L. Lovaśz, Self-dual polytopes and the chromatic number of distance graphs on the sphere. Acta Sci. Math. **45**, 317–323 (1983)
73. V.K. Lyubimov, A.M. Raigorodskii, On lower bounds for the independence numbers of some distance graphs with vertices at $\{-1,0,1\}^n$. Dokl. Math. **80**(N1), 547–549 (2009)
74. F.J. MacWilliams, N.J.A. Sloane, *The Theory of Error-Correcting Codes* (North-Holland, Amsterdam, 1977)
75. M. Mann, Hunting unit-distance graphs in rational n-spaces. Geombinatorics **13**(N2), 49–53 (2003)
76. J. Matoušek, *Using the Borsuk–Ulam Theorem*, Universitext (Springer, Berlin, 2003)
77. N.G. Moshchevitin, A.M. Raigorodskii, On colouring the space $\mathbb{R}^n$ with several forbidden distances. Math. Notes **81**(N5), 656–664 (2007)
78. V.F. Moskva, A.M. Raigorodskii, New lower bounds for the independence numbers of distance graphs with vertices at $\{-1,0,1\}^n$. Math. Notes **89**(N2), 307–308 (2011)
79. O. Nechushtan, Note on the space chromatic number. Discrete Math. **256**, 499–507 (2002)
80. Y. Peres, W. Schlag, *Two Erdős problems on lacunary sequences: chromatic numbers and Diophantine approximations*, Bull. London Math. Soc. **42**(2), 295–300 (2010)
81. R. Radoičić, G. Tóth, *Note on the Chromatic Number of the Space*. Discrete and Computational Geometry: The Goodman–Pollack Festschrift, 695–698. (Algorithms and Combinatorics, 25) (Springer, Berlin, 2003)
82. A.M. Raigorodskii, Borsuk's problem and the chromatic numbers of metric spaces. Russ. Math. Surv. **56**(N1), 103–139 (2001)
83. A.M. Raigorodskii, *The Chromatic Numbers* (Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, 2003) [in Russian]
84. A.M. Raigorodskii, *Borsuk's Problem* (Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, 2006) [in Russian]
85. A.M. Raigorodskii, Around Borsuk's conjecture. J. Math. Sci. **154**(N4), 604–623 (2008)
86. A.M. Raigorodskii, Three lectures on the Borsuk partition problem. Lond. Math. Soc. Lect. Note Ser. **347**, 202–248 (2007)
87. A.M. Raigorodskii, The Borsuk partition problem: the seventieth anniversary. Math. Intel. **26**(N3), 4–12 (2004)
88. A.M. Raigorodskii, On the chromatic number of a space. Russ. Math. Surv. **55**(N2), 351–352 (2000)
89. A.M. Raigorodskii, *The Linear Algebra Method in Combinatorics* (Moscow Centre for Continuous Mathematical Education (MCCME), Moscow, 2007) [in Russian]
90. A.M. Raigorodskii, On the chromatic number of a space with l_q-norm. Russ. Math. Surv. **59**(N5), 973–975 (2004)
91. A.M. Raigorodskii, On the chromatic numbers of spheres in Euclidean spaces. Dokl. Math. **81**(N3), 379–382 (2010)
92. A.M. Raigorodskii, *On the chromatic numbers of spheres in* $\mathbb{R}^n$, Combinatorica, **32** (N1), 111–123 (2012)
93. A.M. Raigorodskii, On distance graphs with large chromatic number but not containing large simplexes. Russ. Math. Surv. **62**(N6), 1224–1225 (2007)
94. A.M. Raigorodskii, The Erdős–Hadwiger problem and the chromatic numbers of finite geometric graphs. Sb. Math. **196**(N1), 115–146 (2005)
95. A.M. Raigorodskii, On a bound in Borsuk's problem. Russ. Math. Surv. **54**(N2), 453–454 (1999)
96. A.M. Raigorodskii, Counterexamples to Borsuk's conjecture on spheres of small radius. Dokl. Math. **82**(N2), 719–721 (2010)
97. A.M. Raigorodskii, On lower bounds for the Borsuk and Hadwiger numbers. Russ. Math. Surv. **59**(N3), 585–586 (2004)
98. A.M. Raigorodskii, On the Borsuk and Erdős – Hadwiger numbers. Math. Notes **79**(N6), 854–863 (2006)
99. A.M. Raigorodskii, M.I. Absalyamova, A lower bound for the chromatic number of the space $\mathbb{R}^n$ with k forbidden distances and metric l_1. Chebyshev Sbornik **7**(N4 (20)), 105–113 (2006) [in Russian]

100. A.M. Raigorodskii, M.M. Kityaev, On a series of questions concerning the Borsuk and the Nelson–Erdős–Hadwiger problems. Math. Notes **84**(N2), 239–255 (2008)
101. A.M. Raigorodskii, A.B. Kupavskiy, Partition of three-dimensional sets into five parts of smaller diameter. Math. Notes **87**(N2), 208–219 (2010)
102. A.M. Raigorodskii, A.B. Kupavskii, M.V. Titova, *On the densest sets without distance 1in spaces of small dimensions*, Trudy of the Moscow Institute of Physics and Technology, **4**, (N1), pp. 111–121 (2012)
103. A.M. Raigorodskii, O.I. Rubanov, in *On the Clique and the Chromatic Numbers of High-Dimensional Distance Graphs*, ed. by S.D. Adhikari, B. Ramakrishnan, Harish-Chandra Research Institute. Number Theory and Applications: Proceedings of the International Conferences on Number Theory and Cryptography (A publication of It's located at Gurgaon, India, 2009), pp. 149–157
104. A.M. Raigorodskii, O.I. Rubanov, On distance graphs with large chromatic number and without large cliques. Math. Notes **87**(N3), 417–428 (2010)
105. A.M. Raigorodskii, I.M. Shitova, On the chromatic numbers of real and rational spaces with several real or rational forbidden distances. Sb. Math. **199**(N4), 579–612 (2008)
106. A.M. Raigorodskii, I.M. Shitova, On the chromatic number of the Euclidean space and on Borsuk's problem. Math. Notes **83**(N4), 579–582 (2008)
107. C. Richter, A. Hinrichs, New sets with large Borsuk numbers. Discrete Math. **270**, 136–146 (2003)
108. C.A. Rogers, Covering a sphere with spheres. Mathematika **10**, 157–164 (1963)
109. C.A. Rogers, Symmetrical sets of constant width and their partitions. Mathematika **18**, 105–111 (1971)
110. O.I. Rubanov, Chromatic numbers of 3-dimensional distance graphs containing no tetrahedra. Math. Notes **82**(N5–6), 718–721 (2007)
111. O. Schramm, Illuminating sets of constant width. Mathematika **35**, 180–189 (1988)
112. G.J. Simmons, On a problem of Erdős concerning 3-colouring of the unit sphere. Discrete Math. **8**, 81–84 (1974)
113. G.J. Simmons, The chromatic number of the sphere. J. Austral. Math. Soc. Ser. **21**, 473–480 (1976)
114. A. Soifer, *The Mathematical Coloring Book* (Springer, Berlin, 2009)
115. L.A. Székely, Erdős on unit distances and the Szemerédi–Trotter theorems, Paul Erdős and his mathematics, bolyai series Budapest. J. Bolyai Math. Soc. **11**, 649–666 (2002)
116. L.A. Székely, N.C. Wormald, Bounds on the measurable chromatic number of $\mathbb{R}^n$. Discrete Math. **75**, 343–372 (1989)
117. D.R. Woodall, Distances realized by sets covering the plane. J. Combin. Theor. A **14**, 187–200 (1973)
118. N. Wormald, A 4-chromatic graph with a special plane drawing. Aust. Mat. Soc. (Ser. A) **28**, 1–8 (1979)
119. A.R. Yarmukhametov, On the connectivity of random distance graphs of a special type. Chebyshev Sbornik **10**(N1 (29)), 95–108 (2009)
120. M.E. Zhukovskii, The weak zero-one laws for the random distance graphs. Dokl. Math. **81**(N1), 51–54 (2010)
121. M.E. Zhukovskii, The weak zero-one law for the random distance graphs. Theor. Prob. Appl. **55**(N2), 344–350 (2010)
122. M.E. Zhukovskii, The weak zero-one law for the random distance graphs. Vestnik Russ. Univ. People's Friendship **2**(N1), 11–25 (2010) [in Russian]
123. G.M. Ziegler, Coloring Hamming graphs, optimal binary codes, and the 0/1-Borsuk problem in low dimensions. Lect. Notes Comput. Sci. **2122**, 159–171 (2001)

Realizability of Graphs and Linkages

Marcus Schaefer

Abstract We show that deciding whether a graph with given edge lengths can be realized by a straight-line drawing has the same complexity as deciding the truth of sentences in the existential theory of the real numbers, ETR; we introduce the class $\exists\mathbb{R}$ that captures the computational complexity of ETR and many other problems. The graph realizability problem remains $\exists\mathbb{R}$-complete if all edges have unit length, which implies that recognizing unit distance graphs is $\exists\mathbb{R}$-complete. We also consider the problem for *linkages*: In a realization of a linkage, vertices are allowed to overlap and lie on the interior of edges. Linkage realizability is $\exists\mathbb{R}$-complete and remains so if all edges have unit length. A linkage is called *rigid* if any slight perturbation of its vertices that does not break the linkage (i.e., keeps edge lengths the same) is the result of a rigid motion of the plane. Testing whether a configuration is not rigid is $\exists\mathbb{R}$-complete.

1 Introduction

Many computational problems in geometry, graph drawing, and other areas can be shown to be decidable using the (existential) theory of the real numbers; this includes the rectilinear crossing number, the Steinitz problem, and finding a Nash equilibrium. What is less often realized—with some exceptions—is that the existential theory of the reals captures the computational complexity of many of these problems precisely: Deciding the truth of a sentence in the existential theory of the reals is polynomial-time equivalent to finding the rectilinear crossing number problem [3], solving the Steinitz problem [4, 34], finding a Nash equilibrium [42], recognizing intersection graphs of convex sets and ellipses [41], recognizing

M. Schaefer (✉)
Department of Computer Science, DePaul University, 243 South Wabash,
Chicago, Illinois 60604, USA
e-mail: mschaefer@cdm.depaul.edu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_24,

unit-disk graphs [32], and many other problems.[1] In this chapter, we try to further substantiate this claim by showing that some well-known Euclidean realizability problems have the same complexity. One consequence of these results is that efficient algorithmic solutions to any of these problems are unlikely, since they would lead to efficient decision procedures for the existential theory of the real numbers, a problem that is **NP**-hard but not known (or expected) to be in **NP**.

1.1 The Existential Theory of the Real Numbers

The existential theory of the real numbers, ETR, is the set of true sentences of the form

$$(\exists x_1,\ldots,x_n)\ \varphi(x_1,\ldots,x_n),$$

where φ is a quantifier-free $(\vee,\wedge,\neg)$-Boolean formula over the signature $(0,1,+,*,<,\leq,=)$ interpreted over the universe of real numbers.[2]

Tarski showed that ETR is decidable, but the running time of his decision procedure is not elementary (that is, bounded above by a tower of exponentials of fixed height).[3] The existential theory of the reals is expressive enough to phrase many interesting problems in robotics and geometry, so research into more practical algorithms for deciding ETR has continued steadily since the 1970s, when Collins discovered cylindrical algebraic decompositions that gave a double exponential time algorithm for deciding ETR. Canny, motivated by problems in robot motion planning, showed that the problem is solvable in **PSPACE** [8]. This is still the best upper bound on ETR in terms of complexity theory, though Renegar sharpened Canny's result in terms of algebraic complexity [36–38]. For a detailed survey, see [33]; for experimental comparisons of running times, see [22].

As it turns out, ETR cannot only be used to solve algorithmic problems, it also captures the complexity of many such problems precisely. To make this statement precise, we use the notion of reducibility from computational complexity: We say a problem A *reduces* to B ($A \leq_{\mathrm{m}} B$) if there is a polynomial-time computable function f so that $x \in A$ if and only if $f(x) \in B$ for all x; the function f is known as the *reduction*.[4] This notion of reducibility is transitive, so we can use it to (partially)

---

[1]A manuscript collecting many of these problems is in preparation (Schaefer, The real logic of drawing graphs, personal communication).

[2]When writing formulas in the existential theory of the reals, we will freely use integers and rationals, since these can easily be eliminated without affecting the length of the formula substantially. We will also drop the symbol $*$.

[3]Tarski showed that the full theory of the reals is decidable by quantifier elimination.

[4]This reducibility is known as polynomial-time many–one reducibility; since we have no need for other reducibilities in this chapter, we simplify to "reduces." We consider decision problems (requiring a "yes" or "no" answer) encoded as sets of binary strings; that is, $A,B \subseteq \{0,1\}^*$, and $f: \{0,1\}^* \to \{0,1\}^*$. For more background on encodings and basic definitions from computational

order problems by their complexity. Intuitively, A reduces to B if A is at most as hard as B, since a solution to B can be combined with the polynomial-time algorithm for f to answer membership in A.[5] We can now define $\exists\mathbb{R}$ to be the complexity class associated with ETR: A decision problem belongs to $\exists\mathbb{R}$ if it reduces to ETR; a decision problem is $\exists\mathbb{R}$-*hard* if every problem in $\exists\mathbb{R}$ reduces to it in polynomial time; it is $\exists\mathbb{R}$-*complete* if it belongs to $\exists\mathbb{R}$ and is $\exists\mathbb{R}$-hard. Analogously, we define notions based on $\forall\mathbb{R}$, the problems whose complement (in $\{0,1\}^*$) is in $\exists\mathbb{R}$. Note that **NP** $\subseteq \exists\mathbb{R}$, since we can express satisfiability of a Boolean formula in $\exists\mathbb{R}$. For example, $(x \vee \bar{y} \vee z) \wedge (\bar{x} \vee y \vee z) \wedge (\bar{x} \vee \bar{y} \vee \bar{z})$ is equivalent to

$$(\exists x, y, z)[x(x-1) = 1 \wedge\ y(y-1) = 1 \wedge\ z(z-1) = 1$$
$$\wedge\ (x(1-y)z) + ((1-x)yz) + ((1-x)(1-y)(1-z)) = 0].$$

We do not know whether **coNP** $\subseteq \exists\mathbb{R}$. By Canny's result, $\exists\mathbb{R} \subseteq$ **PSPACE**, and this is still the known best upper bound on $\exists\mathbb{R}$ [8].

1.2 Stretchability and $\exists\mathbb{R}$-Completeness

To show that a geometric problem is $\exists\mathbb{R}$-complete, we could try reducing from ETR, but that is typically hard. Moreover, it is unnecessary, since there is a prototypical geometric $\exists\mathbb{R}$-complete problem due to Mnëv: stretchability of pseudolines. A *pseudoline* is a simple curve that is *x-monotone*; that is, the curve crosses every vertical line exactly once. An *arrangement of pseudolines* is a finite collection of pseudolines so that any two pseudolines cross exactly once. (We allow multiple pseudolines to cross at the same point.) Two arrangements are *equivalent* if there is a homeomorphism of the plane turning one into the other. An arrangement of pseudolines is *stretchable* if it is equivalent to an arrangement of straight lines. We call the corresponding computational decision problem STRETCHABILITY.

Mnëv showed that ETR reduces to STRETCHABILITY. Shor [43] gives a much simpler proof of this result (also see [39]). In our terminology, we can express Mnëv's result as follows.

Theorem 1.1 (Mnëv [34]). STRETCHABILITY *is* $\exists\mathbb{R}$*-complete.*

Remark 1.2. Mnëv showed a much stronger result, his universality theorem, about the realizability of semialgebraic sets through point-set configurations, of which

---

complexity including complexity classes **NP** (nondeterministic polynomial time) and **PSPACE** (polynomial space), see any of the standard references, e.g., [35, 44]. We could have replaced polynomial time with logarithmic space in the definition of $\leq_{\mathrm{m}}$ but decided to use the more familiar notion.

[5] So "A reduces to B" does not mean that B is easier than A, as the word "reduces" may incorrectly suggest.

Theorem 1.1 is a consequence. We will return to universality briefly in Sect. 2.3 when dealing with issues of precision.

Just because an arrangement is stretchable, it need not be easy to realize; indeed, Mnëv's universality theorem implies that there are stretchable arrangements of pseudolines so that every straight-line realization contains a nonrational line (that is, a line not containing a rational point; an earlier example of this is due to Perles according to Richter-Gebert and Ziegler [18, p.144]). It is known, however, that realizations cannot be arbitrarily complex. This is based on a result by Grigor'ev and Vorobjov [20] on semialgebraic sets. The *total degree* of a (multivariate) polynomial $f : \mathbb{R}^n \to \mathbb{R}$ is the maximum over the sum of variable exponents in each monomial term.

Theorem 1.3 (Grigor'ev, Vorobjov [20, Lemma 9]). *If $f_1, \ldots, f_k : \mathbb{R}^n \to \mathbb{R}$ are polynomials each of total degree at most d and coefficients of bit length at most L, then every connected component of $\{x \in \mathbb{R}^n : f_1(x) \geq 0, \ldots, f_k(x) \geq 0\}$ contains a point of distance less than $2^{Ld^{cn}}$ from the origin, for some absolute constant $c > 0$.*

The theorem (with proof) can also be found in [2, Theorem 13.15].[6] Based on this theorem, Goodman et al. [19] proved the following result. (They phrase the result for point configurations, which are dual to arrangements.)

Lemma 1.4 (Goodman et al. [19]). *A stretchable arrangement of n pseudolines can be realized by n straight lines so that all intersection points lie in the unit disk and so that the distance between any two intersection points and the distance of any intersection point and a line not containing that point is at least $1/2^{2^{cn}}$ for some fixed $c > 0$.*

Goodman et al. also showed the complementary result that some arrangements do require a precision of order $1/2^{2^{cn}}$.

Lemma 1.5 (Goodman et al. [19]). *There are stretchable arrangements of n pseudolines so that any straight-line realization of the arrangement for which all intersection points lie in the unit disk contains two intersections points within a distance less than $1/2^{2^{cn}}$ of each other for some fixed $c > 0$.*

2 Realizability of Graphs

Given a graph $G = (V, E)$ and a *length* $\ell(e) \in \mathbb{R}_{>0}$ for each $e \in E$, is there a straight-line drawing of the graph in the plane (not necessarily crossing-free) where each edge has its prescribed length? If so, we say that the graph is *realizable* in $\mathbb{R}^2$. Realizability depends on the notion of drawing we use; in the standard definition of a drawing, different vertices cannot coincide in the drawing and a vertex cannot

[6] The statement in [2] contains a typo in the radius of the ball.

lie on an edge unless it is an endpoint of that edge.[7] If we do allow vertices to coincide and lie on edges, we enter the realm of linkages. For example, $K_{2,3}$ cannot be realized by a standard straight-line drawing in the plane if all edges have unit length, but it can be realized as a linkage with vertices overlapping.[8] This section will center on graph realizability, while Sect. 3 will discuss linkage realizability.

2.1 Graph Realizability

Theorem 2.1. *Deciding whether a graph with given lengths is realizable is* $\exists\mathbb{R}$*-complete even if all edges have unit length.*

Remark 2.2. We are not aware of any hardness results on the graph (as opposed to the linkage) realizability problem, though David Eppstein writes that he expects that "the Eades–Whitesides logic engine technique can be used to show that it's NP-hard to test whether a graph is a unit distance graph" [14]. There are results on plane realizations, which we survey in Sect. 2.4. The special case of the complete graph turns out to be efficiently solvable: Lemke, Skiena, and Smith sketch an algorithm that shows how to determine realizability of the complete graph and compute the coordinates of the points [29]. The history of complexity results on linkage realizability is discussed in Remark 3.2.

In the proof of Theorem 2.1, we make use of the Peaucellier linkage, a beautiful device transforming circular motion into linear motion.[9] Figure 1 shows the linkage;

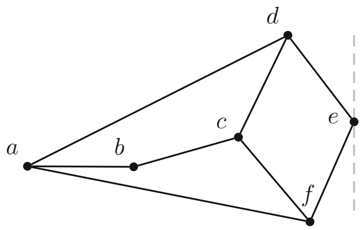


Fig. 1 A Peaucellier linkage

[7]For a discussion of graph drawing assumptions, see [47].

[8]Our distinction between the realizability of graphs and linkages is not universal, but not unusual; for linkages, [10] is a good reference; for graph realizability, definitions and terminology vary. For instance, realizable graphs are sometimes called *Euclidean graphs*. Often the definitions allow that vertices lie on edges of which they are not endpoints, though that may in some cases be due to oversight; as we will see, from a computational point of view, this distinction does not matter.

[9]There are many applets implementing Peaucellier's linkage available on the web to play with [30]. James Joseph Sylvester writes about Lord Kelvin that he "nursed it as if it had been his own child, and when a motion was made to relieve him of it, replied 'No! I have not had nearly enough of it—it is the most beautiful thing I have ever seen in my life'." [46]. There have been other devices, both earlier and later, achieving the same effect, see [26] for an early history.

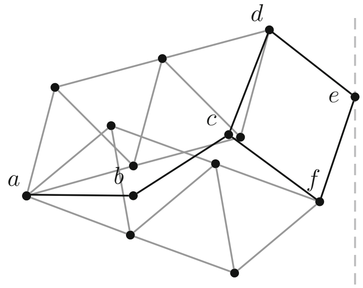


Fig. 2 A Peaucellier linkage with edges of unit length

we require $|ab| = |bc|$, $|ad| = |af|$ and $|cd| = |de| = |ef| = |fc|$. If we keep points a and b fixed, and move c (on a circular arc with center b), then e moves on a straight line. In other words, the *locus* of e is a straight-line segment [21].

Figure 2 shows how to construct this Peaucellier linkage as a realizable graph with edges of unit length; we call this gadget $P_{a,b}(e)$ (edges ad and af have been replaced by two rigid parts with unnamed vertices).

Note that in the drawing all vertices have distinct positions, and no vertex lies on an edge without being an endpoint of that edge. By continuity, this remains true if we move e slightly from its initial position. In general, we will need to place the Peaucellier gadget into partially completed drawings, and at that point we need to keep ensuring the basic graph drawing conventions. We call a drawing *nice* if (1) no vertex lies on a line of which it is not the endpoint, (2) no more than two edges cross in a point, (3) two vertices that are not adjacent have a distance different from 1.

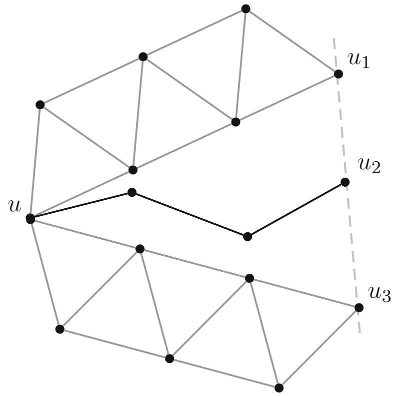
With the Peaucellier gadget, we can build a "colinearity" gadget that ensures that three points lie on a line.

Lemma 2.3. *There is a gadget $C(u_1,u_2,u_3)$ with edges of unit length so that in any realization of $C(u_1,u_2,u_3)$ the points u_1, u_2, and u_3 lie on a line. On the other hand, if three distinct points u_1, u_2, and u_3 lie on a line segment of length at most $\delta > 0$, for some fixed δ, then $C(u_1,u_2,u_3)$ is realizable. Moreover, $C(u_1,u_2,u_3)$ can be added to an existing nice drawing with three such vertices u_1, u_2, u_3 so that the resulting drawing remains nice.*

Proof. We will use Peaucellier's linkage to guarantee colinearity: Given three vertices $(u_i)_{i\in[3]}$, create three copies $(P_i)_{i\in[3]}$ of P, identify $a := a_1 = a_2 = a_3$ and $b := b_1 = b_2 = b_3$, and u_i with e_i, $i \in [3]$ to obtain $C_{a,b}(u_1,u_2,u_3)$. The realizability of $C_{a,b}(u_1,u_2,u_3)$ guarantees that u_1, u_2, and u_3 are colinear by the properties of the Peaucellier gadget.

For the reverse direction, assume we start with a nice drawing containing three vertices u_1, u_2, u_3 lying on a sufficiently short line segment. Pick a and b so that $C_{a,b}(u_1,u_2,u_3)$ is realizable and nice (by itself) and remains so for small perturbations of ab. The resulting drawing may not be nice, since two nonadjacent vertices may have a distance of 1, two distinct vertices overlap, or a vertex lies on an edge. Perturbing ab slightly can destroy all three of these events, since it will perturb the locations of all the vertices in $C_{a,b}(u_1,u_2,u_3)$; the only pairs of vertices whose distance remains constant have an edge between them. □

Fig. 3 $B(u_1, u_2, u_3)$



To encode STRETCHABILITY we need to express that a point u_2 lies between u_1 and u_3. For this we need another gadget, $B(u_1, u_2, u_3)$.

Lemma 2.4. *There is a gadget $B(u_1, u_2, u_3)$ with edges of unit length so that in any realization of $B(u_1, u_2, u_3)$ in which u_1, u_2, and u_3 lie on a line, u_2 lies between u_1 and u_3. On the other hand, if u_2 lies on the line segment u_1u_3 and u_1u_3 has length at most $\delta > 0$, then $B(u_1, u_2, u_3)$ is realizable.*

Proof. Consider the graph $B(u_1, u_2, u_3)$ shown in Fig. 3. In any realization of that graph, uu_1 and uu_3 have distance 3, while uu_2 has distance at most 3; hence, if all three points lie on the same line, then u_2 must lie between u_1 and u_2. □

We can now complete the proof of Theorem 2.1.

Proof of Theorem 2.1. It is easy to see that the problem lies in $\exists\mathbb{R}$ (note that we do not have to calculate any square roots). We reduce from STRETCHABILITY: Suppose we are given a pseudoline arrangement $\mathcal{A}$. Create a vertex for every intersection point. For any three consecutive points u_1, u_2, u_3 along a pseudoline, add the devices $C(u_1, u_2, u_3)$ and $B(u_1, u_2, u_3)$. Call the resulting graph $G_{\mathcal{A}}$ (all edges having unit length). If $G_{\mathcal{A}}$ is realizable, then a realization contains a set of line segments whose order of intersection corresponds to $\mathcal{A}$. Since every two of these lines intersect (we included every intersection point in $G_{\mathcal{A}}$), we can extend these line segments to infinite straight lines without changing the order type; hence, $\mathcal{A}$ is stretchable.

In the reverse direction, we have to show that if $\mathcal{A}$ is stretchable, then $G_{\mathcal{A}}$ is realizable. This would be entirely straightforward if we did not have to ensure that the drawing is nice. Let's assume that $\mathcal{A}$ is stretchable. Then there is a realization by straight lines in which all intersections lie within the unit disk centered at the origin (using dilation). Let D_δ be the drawing containing all intersection points dilated by a factor of $\delta > 0$, where δ is chosen sufficiently small for the B and C gadgets to work properly. Note that D_δ is a nice drawing, trivially so, since all points have distance less than 1 and there are no edges yet. To this drawing add the two rigid legs of each B gadget, obtaining a drawing D'_δ. The resulting drawing may not necessarily be nice, since there could be two nonadjacent vertices that have a distance of 1, two vertices that overlap, or a vertex lying on an edge.

Let p be a point on one of the gadgets B (p does not have to be a vertex); we write $p(\delta)$, thinking of p as depending on δ. The coordinates of $p(\delta)$ are nearly polynomials in δ, except that they may contain a term of the form $\sqrt{1-s\delta}$ for some fixed $s \in \mathbb{R}$ only depending on the placement of the B gadget that p belongs to in the initial drawing D. The square of the distance between two points (not necessarily belonging to the same B), or the square of the distance between a point and a line containing an edge (not necessarily belonging to the same gadget as the point) can similarly be expressed as a near-polynomial term in δ with at most three terms $\sqrt{1-s_i\delta}$, $i = 1,2,3$. Hence, the condition that such a distance is 0 or 1 can be expressed using an equation that is nearly polynomial, except for at most three terms of the form $\sqrt{1-s_i\delta}$. However, three such terms can be removed (using symmetric polynomials), replacing the distance conditions with purely polynomial equations, which either have a finite number of solutions or are true for all δ. As we let δ go to zero, the different B gadgets converge to different locations (since the lines they are based on cannot be parallel: Every pair of lines crosses), so it is not possible that two points belonging to different B gadgets always have distance 0 or 1, and similarly a point in one B gadget cannot always have distance 0 from a line in another B gadget. Finally, if two points within the same B gadget always have distance 1, they have an edge between them, and the distance between a point and a line in the same B gadget changes with δ, unless the point is an end vertex of an edge of B. These observations imply that there are only a finite number of values of δ for which D'_δ is not nice. Hence, we can pick an arbitrarily small $\delta > 0$ for which the drawing D'_δ is nice. Fix such a δ.

We can now add the flexible middle leg of each B gadget, so that the additional vertices maintain niceness (this is why the middle leg has two interior vertices: to be flexible enough to maintain niceness).

Finally, Lemma 2.3 allows us to add all the C gadgets one by one, maintaining niceness of the drawing. This shows that $G_{\mathcal{A}}$ is realizable, even fulfilling the stronger condition that every pair of nonadjacent vertices has a distance different from 1. □

Corollary 2.5. *The problem of graph realizability remains* $\exists\mathbb{R}$*-complete even if* (i) *we do not require that vertices not lie on edges they are not endpoints of and* (ii) *nonadjacent vertices must have a distance different from* 1.

Proof. This follows from the proof of Theorem 2.1: The proof never used assumption (i), and when constructing the realization of $G_{\mathcal{A}}$, we ensured that nonadjacent vertices have distance different from 1. □

2.2 Euclidean Dimension and Unit Distance Graphs

Let E^n be the infinite graph on vertex set $\mathbb{R}^n$ so that xy is an edge of E^n if and only if $|x-y|=1$. Then Corollary 2.5 implies that recognizing subgraphs or induced subgraphs of E^2 is $\exists\mathbb{R}$-complete (independent of how we arbitrate the issue of vertices lying on edges). The *Euclidean dimension* of a graph G is the smallest n so that G is an (induced) subgraph of E^n, a notion introduced by Erdős et al. [16].[10]

Corollary 2.6. *Deciding whether a graph has Euclidean dimension* 2 *is* $\exists\mathbb{R}$-*complete (in both the induced and noninduced version). Hence, computing the Euclidean dimension of a graph is* $\exists\mathbb{R}$-*complete.*

Subgraphs of E^2 are also known as *unit distance graphs* [9], while *strict unit distance graphs* are induced subgraphs of E^2. The following is just a restatement of the preceding corollary, answering a question suggested by Eppstein [14].

Corollary 2.7. *Recognizing (strict) unit distance graphs is* $\exists\mathbb{R}$-*complete.*

Consequently, it is very unlikely that we will be able to recognize unit distance graphs efficiently [24].

Remark 2.8. We can think of unit distance graphs as graphs whose edges are labeled "$=1$"; this suggests looking at alternative label sets. For example, if edges can have labels "<1" and ">1" instead of "$=1$," we get *unit-disk graphs*; McDiarmid and Müller [32] recently showed that SIMPLE STRETCHABILITY (in which no more than two pseudolines are allowed to intersect in a point) reduces to the recognition problem for unit-disk graphs, so the unit-disk graph problem is also $\exists\mathbb{R}$-complete, since SIMPLE STRETCHABILITY and STRETCHABILITY are polynomial-time equivalent [42].

2.3 Issues of Precision

The reduction from STRETCHABILITY to graph realizability in the proof of Theorem 2.1 is *geometric* in the following sense: Each realization of the graph with length constraints constructed from the pseudo-line arrangement encodes a straight-line realization of the arrangement. More precisely, the position of certain vertices of the graph encodes the locations of intersection points of the straight-line realization. In other words, the intersection points of a realization can be obtained by projecting

[10] Erdős et al. [16] defined the *dimension* of a graph using (noninduced) subgraphs of E^n. Later, Erdős and Simonovits [15] introduced Euclidean dimension under the name *faithful dimension*. The two notions differ: Take a wheel W_6 with six spokes and remove one of the spokes. The resulting graph is realizable as a subgraph, but not as an induced subgraph of E^2. The name "Euclidean dimension" seems to be due to Maehara [31]. For details and more terminology and history, see [45, Sect. 13.2].

onto certain points of the graph. Lemma 1.5 now immediately implies that some graph realizations have exponentially low vertex resolution: The ratio between the maximum distance of any two vertices divided by the minimum distance between any two (distinct) vertices can be of order $2^{2^{cn}}$. Similarly, the reductions to Euclidean dimension and unit distance graphs are geometric. We state the result for unit distance graphs only.

Corollary 2.9. *There are unit distance graphs on n vertices, so that any realization of the graph contains two distinct vertices at distance at most* $1/2^{2^{cn}}$ *for some fixed constant* $c > 0$.

Some of the traditional ETR results are actually *universality theorem*; for example, Mnëv [34] showed that any semialgebraic set is *stably equivalent* to the realization space of a pseudoline arrangement. We do not want to define stable equivalence explicitly (see [39] for a detailed discussion), but roughly speaking it means that the two sets look very similar algebraically. Stable equivalence is not immediately useful to our purposes, since it does not imply any complexity bounds, but many of the universality theorems in the literature could be recast as polynomial-time many–one reductions. We also think it likely that many of our geometric reductions can be turned into universality theorems with some additional effort. We discuss universality theorems for linkages in Remark 3.3.

2.4 Plane Realizations and Matchstick Graphs

It is natural to ask what happens if we require the realizations of the graphs to be plane, that is, free of crossings. Strengthening earlier results by Whitesides [48] and Eades and Wormald [13], Cabello et al. [7] showed that the plane realizability problem is (strongly) **NP**-hard even when restricted to 3-connected, infinitesimally rigid planar graphs with unit edge lengths. Plane realizable graphs with unit edge lengths—plane unit distance graphs—are known as *matchstick graphs*.

It remains open whether recognizing matchstick graphs, or solving the general plane realizability problem, is $\exists\mathbb{R}$-complete. If the graph is a triangulation, then Cabello, Demaine, and Rote [7] show that plane realizability can be decided in polynomial time in the real RAM model, and in **P** if the graph has bounded degree. Their proof can be modified to show that the problem lies in **coRP** without assuming bounded degree; a problem lies in **coRP** if there is a probabilistic algorithm that is correct if it says "no" and is correct with probability at least $1/2$ if it says "yes" [35, 44]. Running the algorithm repeatedly yields an algorithm with an arbitrarily small error bound.

Lemma 2.10. *If G is a triangulation with prescribed lengths, then plane realizability can be tested in* **coRP***, even if the edge lengths include square roots of rationals.*

Our proof adapts an argument from Cabello et al. [7].

Proof. By a result of Di Battista and Vismara [12], it is sufficient to verify the triangle inequality in each triangle, and ensure that the sum of angles at each interior vertex is 2π. Since the graph is a triangulation, we can easily determine its topological embedding (if there is a nontriangle face, it has to be the outer face; if all faces are triangles, we can try each of them as the outer face in polynomial time). Now pick an interior vertex v and consider one of its incident triangles; the angle α formed by the triangle at v fulfills the law of cosines: $\cos\alpha = (b^2+c^2-a^2)/2ab$, with standard notation for the triangle. Consequently, $\sin\alpha = \sqrt{1-((b^2+c^2-a^2)/2ab)^2}$. Suppose the angles at v are α_i, $i \in [n]$, and let $A_k := \sum_{i=1}^{k} \alpha_i$. We can then write recursive expressions for $\sin A_k$ and $\cos A_k$ using identities $\sin(\alpha+\beta) = \sin\alpha\cos\beta + \cos\alpha\sin\beta$ and $\cos(\alpha+\beta) = \cos\alpha\cos\beta + \sin\alpha\sin\beta$. This expression can be viewed as a directed acyclic graph whose leaf nodes are the only nodes containing the square root operation (note that this remains true even if the input lengths are square roots of rationals). Blömer [5, Theorem 2.2] showed that deciding whether such an expression equals 0 lies in **coRP**. In particular, testing whether $\sin A_n = 0$ and $\cos A_n = 1$ is in **coRP**. We also need to ensure that all $A_k < 2\pi$, but it is sufficient to do this approximately, since the test of $\sin A_n = 0$ and $\cos A_n = 1$ guarantees that A_n is a multiple of 2π, and the approximate test can be done in polynomial time. □

If recognizing matchstick graphs were $\exists\mathbb{R}$-complete and the reduction were geometric, then, similar to Corollary 2.9, we would need to be able to construct matchstick graphs in which vertices are forced to be exponentially close. This would be a first indication of potential $\exists\mathbb{R}$-hardness.

Question 2.11. Can one construct a matchstick graph on n vertices so that in every plane realization of the graph, there are two vertices of distance at most $1/2^{2^{cn}}$ for some fixed constant c?

3 Realizability of Linkages

A *linkage* is a graph $G = (V,E)$ with a *length* $\ell(e)$ assigned to each edge $e \in E$; edges of a linkage are also called *rods* or *bars*; a *configuration*, or *realization*, of G in $\mathbb{R}^2$ is a mapping of V to $\mathbb{R}^2$ so that the distance between the endpoints of each edge e equals $\ell(e)$. If there is a configuration of the linkage, we call the linkage *realizable*; our terminology is based on the exposition in [10]; a more detailed treatment can be found in [11]. We discuss two computational problems related to linkages: realizability, in Sect. 3.1; and rigidity, in Sect. 3.2.

3.1 Linkage Realizability

Theorem 3.1. *Deciding whether a linkage is realizable is $\exists\mathbb{R}$-complete even if all edges have unit length.*

The $\exists\mathbb{R}$-hardness of linkage realizability can also be obtained from universality results on linkages; see the discussion in Remark 3.3. The main new contribution in Theorem 3.1 is the restriction to unit lengths(and a simpler proof).

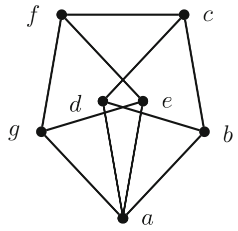
Remark 3.2. The **NP**-hardness of linkage realizability was shown by Yemini [49] and Saxe [40]. Saxe showed that linkage realizability in $\mathbb{R}$ is **NP**-complete even if distances are restricted to values $1,2$: Let $\mathbf{w} := \{w_1,\ldots,w_n\}$ be an instance of the partition problem, and create the graph C_n on edges $e_1,\ldots,e_n$ with $\ell(e_i) = w_i$. Then C_n can be realized in $\mathbb{R}$ if and only if $\mathbf{w}$ is a positive instance of partition. Using gadgets with edges ab, bc, and ac, where $w(ab) = w(bc) = 1$ and $w(ac) = 2$, we can replace edges of arbitrary integer length with edges of lengths 1 and 2. Saxe extended this construction to show that linkage realizability in $\mathbb{R}^k$ for linkages with distances $1,2$ is **NP**-hard for all $k \geq 1$. He also discusses approximation results and results on how hard it is to decide whether there is a unique solution. Abbott [1] showed that deciding whether a linkage is rigid is hard; we discuss his result later in this section.

In the algebraic community, the linkage realizability problem is known as the *Euclidean distance matrix completion problem (EDMCP)*. Research in that community seems to concentrate on algebraic characterizations and using tools such as semidefinite programming to solve instances of the problem; see [28] for a survey.

Remark 3.3. Kempe claimed his universality theorem for linkages in 1876. Roughly speaking, it shows how to trace any (compact) algebraic curve with a linkage, so linkages are universal for (compact) algebraic curves.[11] This suggests a proof of $\exists\mathbb{R}$-hardness for linkage realizability: to test whether a polynomial f can take on the value 0 and extend the linkage tracing f so that the vertex tracing f is forced to lie on the line corresponding to $f = 0$. Then $f = 0$ has a solution if and only if the extended linkage is realizable. This approach can be made to work, but there are some obstacles to overcome. First, Kempe's proof is incomplete (some of his gadgets have degenerate configurations). Jordan and Steiner [23] and Kapovich and Millson [25] gave the first correct and complete proofs of Kempe's universality theorem (and in Kapovich and Millson's case, even stronger results). The next obstacle is that these papers did not analyze the effectiveness of the constructions; Gao and Zhu [17] analyzed Kempe's proof and showed that in the plane, $O(n^4)$ links are sufficient (also see Abbott [1, Sect. 1.5]), this still does not show that the lengths of the links can be calculated effectively (or can be assumed to be algebraic or rational numbers). The first detailed analysis of this aspect seems to be in Abbott's thesis [1], which we will discuss again below with respect to rigidity. Abbott's version of Kempe's universality theorem can be extended to a proof of $\exists\mathbb{R}$-hardness of linkage realizability as sketched above. However, the approach is quite intricate (after all, a much stronger result is obtained: universality); our proof

[11] See [25] or [11, Sect. 3.2.1] for detailed expositions.

Fig. 4 The Moser graph



below is more direct, and, moreover, it shows that unit lengths are sufficient to get ∃ℝ-hardness, while, as far as I know, there are no universality theorems for unit linkages.

The proof of Theorem 3.1 requires some modifications to the proof of Theorem 2.1; in some respects, the proof becomes easier, since we no longer have to ensure that vertices do not accidentally overlap with other vertices or edges in realizations. On the other hand, we need a new device that guarantees that vertices of linkages are mapped to distinct points of the plane.

In the construction, we will use a small set of radii (the particular set of radii is rather arbitrary), so we first show that if we are given a linkage with multiple integer lengths, we can replace it with a unit-length linkage without affecting realizability.

Lemma 3.4. *Given a linkage G with integer lengths, we can construct a linkage G' with unit lengths only, so that G is realizable if and only if G' is realizable. The size of G' is polynomial in the size of G and the integer lengths (in unary).*

Proof. Let M be the *Moser spindle* shown in Fig. 4.

The seven vertices of M are distinct in any realization of M in the plane: Obviously, the three vertices in each of the four triangles are distinct. Now the distance between a and c is either 0 or $\sqrt{3}$, and the same is true for a and f. So it is not possible that one of c or f or both of them coincide with a, since c and f have a distance of 1. Hence, both c and f are distinct from a, and this already forces all vertices of M to be pairwise distinct. In particular, the diamond on $\{a,b,c,d\}$ does not collapse. We can thus chain several Moser graphs together by identifying them along triangles to create pairs of vertices at arbitrary integer distances. This allows us to replace an edge of length n in G using a gadget with less than $7(2n-1)$ vertices. □

Lemma 3.4 helps us resolve the problem of keeping vertices distinct: We cannot simply paste two triangles together along an edge, since the resulting diamond linkage can be realized by a single triangle, with two of the vertices (and two pairs of edges) coinciding. With the Moser graph, we can avoid collapses of this type. For the proof of Theorem 3.1, we need devices that keep arbitrary pairs of vertices apart in a realization. This problem comes in two types: vertices like the two vertices in the diamond that are either identical or far apart, and vertices that could potentially be very close in a realization, but are not allowed to coincide. The second problem is more difficult, and we will show how to resolve it presently. The first problem only occurs in the construction of the Peaucellier linkage, where we resolve it ad hoc.

Lemma 3.5. *We can create a linkage P' with $V(P')$ including a, b, and e so that if a and b are fixed in the plane, then the locus of e is a line segment of length at least* 1.

Proof. We start with the Peaucellier linkage P shown in Fig. 2 assigning each edge a length of 2. In a linkage realization of P several pairs of points that need to be distinct for the gadget to work can collapse; this includes the diamonds connecting a to d and a to f as well as the pairs d, f and c, e. Suppose x and y is one of these pairs; then we add a rather crude device to the linkage: Add edges xx', $x'y'$, and $y'y$ of lengths $w(xx') = 1$, $w(x'y') = 3$, and $w(y'y) = 1$. In any realization, this forces xy to have distance at least 1 (and x and y can have distance up to 5, which is sufficient for all realizations of P). The resulting linkage P' fulfills the statement of the lemma. □

With the Peaucellier linkage, we can now construct a linkage $L(u_1,u_2,u_3)$ that combines the functionality of the earlier colinearity and betweenness gadgets $B(u_1,u_2,u_3)$ and $C(u_1,u_2,u_3)$; as opposed to these earlier gadgets, the new device does not guarantee that u_2 is *strictly* between u_1 and u_3: u_2 could coincide with either. This is a problem we fix later.

Lemma 3.6. *There is a linkage $L(u_1,u_2,u_3)$ with lengths in* $[4]$ *so that the realizability of $L(u_1,u_2,u_3)$ implies that u_2 lies on the line segment u_1u_3 (including endpoints), while, in the reverse direction, if u_2 does lie on the line segment u_1u_3 and u_1u_3 has length at most* 1*, then $L(u_1,u_2,u_3)$ is realizable.*

Proof. Given three vertices $(u_i)_{i\in[3]}$, create three copies $(P_i')_{i\in[3]}$ of P' as constructed in Lemma 3.5, identify $a := a_1 = a_2 = a_3$ and $b := b_1 = b_2 = b_3$, and u_i with e_i, $i \in [3]$. Realizability of this part guarantees that u_1,u_2,u_3 are colinear. To ensure that u_2 lies between u_1 and u_3, we add two new vertices g,h and edges gh, hu_2, gu_1, gu_3 with $w(gh) = w(hu_2) = 2$ and $w(gu_1) = w(gu_3) = 4$. The resulting device is $L(u_1,u_2,u_3)$. If $L(u_1,u_2,u_3)$ is realizable, then u_2 lies on u_1u_3 (including endpoints), and if u_1u_3 has length at most 1 and u_2 lies on that segment, then $L(u_1,u_2,u_3)$ is realizable. □

Finally, we need a gadget $D(u_1,u_2)$ that guarantees that u_1 and u_2 are distinct, where u_1 and u_2 are two intersection points. In general, such a gadget does not exist: If the loci of u_1 and u_2 can come arbitrarily close, then they must intersect, since both are compact sets. However, for realizations of stretchable arrangements, we have Lemma 1.4, which gives us a lower bound on how close intersection points need to get, and we can use that to build a device that simulates true "distinctness" well enough to work for our construction.

Lemma 3.7. *There is a linkage $T(a,b)$ with lengths in* $[2]$ *and $\{a,b,c\} \subset T(a,b)$ so that if we assign a and b to any two points of distance less than* 1 *in the plane, then a and c have distance $|a-b|^2$ in any realization of $T(a,b)$.*

The gadget is based on the von Staudt constructions also used by Mnëv [34].

Proof. We first build a gadget that ensures that two lines are parallel.[12] Take a C_6 on vertices $p_1,\dots,p_6$ and add edge p_2p_5. Let all these edges have unit length, and add two edges p_1p_3 and p_4p_6 of length 2. In any realization of this graph, the lines through p_1p_2 and p_4p_5 are parallel, and p_1/p_5 and p_2/p_4 can get arbitrarily close together (even coincide). Now add Peaucellier linkages $L(p_1,a,p_2)$, $L(p_1,b,p_2)$, $L(p_4,c,p_5)$, $L(p_4,d,p_5)$. The resulting gadget is $P(a,b,c,d)$; note that ab and cd are parallel in any realization of $P(a,b,c,d)$, and if we are given points $\{a,b,c,d\}$ so that ab and cd are parallel, and all points in $\{a,b,c,d\}$ lie within a unit disk, then $P(a,b,c,d)$ is realizable.

To build $T(a,b)$, start with vertices a,b, add vertices c,u,v,w and edges au, av, uv of unit length, add two Peaucellier linkages $L(a,b,u)$ and $L(a,c,v)$, and add $P(u,v,b,w)$ and $P(u,w,b,c)$. Then $|a-c|$ fulfills $|a-c|/|a-w| = |a-b|/|a-u| = |a-b|$, so $|a-c| = |a-w|\cdot|a-b| = |a-b|^2$, since $|a-w| = |a-b|$, as uv and bw are parallel. Furthermore, if a and b have distance at most 1, then $T(a,b)$ is realizable. □

Corollary 3.8. *We can build a linkage $D(u_1,u_2)$ so that u_1,u_2 are distinct in any realization of $D(u_1,u_2)$; moreover, for any u_1, u_2 that have distance at least $1/2^{2^{n^2}}$ (and distance at most 1), there is a realization of $D(u_1,u_2)$. The linkage $D(u_1,u_2)$ has size at most polynomial in n.*

Proof. We take n^2 copies of $T(a,b)$ from Lemma 3.7 and chain them together by identifying a,c with a,b of the next device. Finally, we identify the vertices a,b of the first $T(a,b)$ with the vertices d,e of the Moser spindle (Fig. 4) so that a and b have distance less than $1/2$ (indeed, closer to 0.47).[13] Then, by the properties of $T(a,b)$, the vertices a,c of the last copy of $T(a,b)$ have distance $|a-c| = |d-e|^{2^{n^2}} < 1/2^{2^{n^2}}$.

Given vertices u_1,u_2, take two copies of L as constructed in Lemma 3.6 on vertices $L(u_1,a,b)$ and $L(a,b,u_2)$. The resulting device $D(u_1,u_2)$ forces u_1 and u_2 to have distance at least $|d-e|^{2^{n^2}} > 0$ and are thus distinct. Also, for any u_1,u_2 that have at least distance $1/2^{2^{n^2}}$, gadget $D(u_1,u_2)$ is realizable. □

We are finally in a position to prove Theorem 3.1.

Proof (Proof of Theorem 3.1). It is easy to see that the problem lies in $\exists\mathbb{R}$. We reduce from STRETCHABILITY.

---

[12] Kempe used a parallelism gadget like this in his proof of the universality theorem that every bounded part of an algebraic curve can be traced by a suitable linkage. His parallelism gadget was flawed, however; there are many ways to repair the construction. We follow a construction due to Kapovich and Millson [25], also described in [11, Sect. 3.2.2].

[13] The realization of the Moser graph is not unique; both of the diamonds can flip; to force d and e to be realized at distance $< 1/2$ as shown in Fig. 4, we brace the construction by adding edges gx, xy, and yb of lengths $w(gx) = 1$, $w(xy) = 3$, $w(yb) = 1$; this forces g and b to have distance at least 1, thereby forcing the intended realization of the Moser graph.

Suppose we are given a pseudoline arrangement $\mathcal{A}$. Create a vertex for every intersection point. For any three consecutive points u_1, u_2, u_3 along a pseudoline, add the device $L(u_1, u_2, u_3)$. For any two intersection points u_1, u_2, add the device $D(u_1, u_2)$. By Lemma 3.4, we can assume that the resulting graph $G_{\mathcal{A}}$ has edges of unit length only. If $\mathcal{A}$ is stretchable, then there is a realization of $\mathcal{A}$ by straight lines in which all intersections lie within the unit disk and any two intersection points have distance at least $1/2^{2^{cn}}$ for some fixed $c > 0$ by Lemma 1.4. But then $G_{\mathcal{A}}$ is realizable as long as $n \geq c$, because then $1/2^{2^{n^2}} < 1/2^{2^{cn}}$. On the other hand, if we assume that $G_{\mathcal{A}}$ is realizable, then a realization contains a set of line segments whose order types correspond to $\mathcal{A}$. Since every two of these lines intersect (we included every intersection point in $G_{\mathcal{A}}$), we can extend these line segments to infinite straight lines without changing the order type, and hence $\mathcal{A}$ is stretchable. □

3.2 *Rigidity*

Two configurations $p, p' : V \to \mathbb{R}^2$ of the same linkage $G = (V, E, w)$ are *congruent* if $|p(u) - p(v)| = |p'(u) - p'(v)|$ for all pairs $u, v \in V$. A configuration $p : V \to \mathbb{R}^2$ of $G = (V, E, w)$ is *rigid* if there is a $\varepsilon > 0$ so that any configuration $p' : V \to \mathbb{R}^2$ that is close to p in the sense that $|p(v) - p'(v)| < \varepsilon$ for all $v \in V$ is congruent to p. Informally speaking, the configuration cannot be changed by perturbing points slightly.

Abbott [1] showed that rigidity is **coNP**-hard using the following argument: Let ISO be the problem of deciding whether a family f_i, $i \in [s]$, of polynomials in n variables has an isolated zero. Let $\mathsf{H_2N}$ be the following computational problem: Given a family f_i, $i \in [s]$, of s homogeneous polynomials in n variables, is there a nontrivial zero, i.e., $(x_1, \ldots, x_n) \neq \mathbf{0}$, so that $f_i(x_1, \ldots, x_n) = 0$ for all $i \in [s]$?[14] Then $\mathsf{H_2N}$ reduces to $\overline{\mathsf{ISO}}$ based on the observation that a family of homogeneous polynomials has a nontrivial zero if and only if $\mathbf{0}$ is not an isolated zero of this family [1, Corollary 5.6]. Moreover, ISO reduces to rigidity, a nontrivial reduction due to Abbott [1, Theorem 5.7], and part of his thesis on Kempe's universality theorem; the reduction is in polynomial time, as long as the total degree of the polynomials is bounded by a constant. Together with Koiran's result that $\mathsf{H_2N}$ is **NP**-hard even for polynomials of total degree 2, this implies Abbott's result that rigidity is **coNP**-hard. However, as we are about to show, $\mathsf{H_2N}$ is not only **NP**-hard, but $\exists\mathbb{R}$-complete, so that rigidity is $\forall\mathbb{R}$-hard using the same chain of reductions that Abbott uses. The following lemma is a stepping stone to the $\exists\mathbb{R}$-hardness of $\mathsf{H_2N}$.

Lemma 3.9. *Deciding whether a family of polynomials $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in [s]$ has a common root in $B^n(\mathbf{0}, 1)$ (the unit ball) is $\exists\mathbb{R}$-complete. We can assume that all f_i have total degree at most* 2.

[14] The name $\mathsf{H_2N}$ seems to be short for Hilbert's homogeneous Nullstellensatz [27].

Proof. It is well known that deciding whether a family of polynomials $g_i : \mathbb{R}^m \to \mathbb{R}$, $i \in [s]$, has a common root is $\exists\mathbb{R}$-complete [42], even if all g_i have total degree at most 2.[15] We want to reduce this problem to the problem of deciding whether a family of polynomials f_i has a common root in $B^n(\mathbf{0}, 1)$.

Suppose we are given a family of polynomials $g_i : \mathbb{R}^m \to \mathbb{R}$, $i \in [s]$ of total degree at most 2. By Theorem 1.3, we know that if the g_i, $i \in [s]$, have a common root, then such a root has distance less than $R = 2^{L2^{cn}}$ from the origin, where L is an upper bound on the bit lengths of the coefficients of the g_i and $c > 0$ is some fixed constant (we use that the g_i have total degree 2). Let $t = \lceil n \log c + \log L \rceil + 1$ (so $2^{2^t} > R^2$) and define

$$f_i(x_1, \dots, x_m, y_0, y_1, \dots, y_t) := y_0^2 g_i(x_1/y_0, \dots, x_m/y_0)$$

for $i \in [s]$ and

$$f_{s+1}(x_1, \dots, x_m, y_0, y_1, \dots, y_t) := y_t - 1/2$$

and

$$f_{s+1+i}(x_1, \dots, x_m, y_0, y_1, \dots, y_t) := y_{i-1} - y_i^2$$

for $i \in [t]$. Note that all the f_i have total degree at most 2.

If the f_i, $i \in [s+1+t]$ have a common root $(x_1, \dots, x_m, y_0, y_1, \dots, y_t)$, then $y_t = 1/2$, since $f_{s+1} = 0$, and $y_{t-i} = 2^{-2^i}$ for $i \in [t]$ since $f_{s+i+1} = 0$ for $i \in [t]$. In particular, $y_0 > 0$, so $f_i = 0$ implies that $g_i(x_1/y_0, \dots, x_m/y_0) = 0$, for all $i \in [s]$, so $(x_1/y_0, \dots, x_m/y_0)$ is a common root of the g_i, $i \in [s]$.

On the other hand, assume that the g_i have a common root $(x'_1, \dots, x'_m)$. We can assume that $(x'_1, \dots, x'_m)$ has distance less than R from the origin. Let $y_i := 2^{-2^{t-i}}$ for $i \in \{0\} \cup [t]$ and $x_i := x'_i y_0$. By definition, all $f_i = 0$, $i \in [s+1+t]$. We only need to verify that $(x_1, \dots, x_m, y_0, y_1, \dots, y_t) \in B^{m+1+t}(\mathbf{0}, 1)$. Now $\sum_{i=0}^t y_i^2 \leq \sum_{i=0}^\infty 4^{-2^i} = 1/4 + 1/16 + 1/256 + 1/65536 + \cdots \leq 1/4 + 1/8 = 5/8$. Also, $\sum_{i=1}^m x_i^2 = y_0^2 \sum_{i=1}^m x_i'^2 \leq y_0^2 R^2 \leq (1/R^2)^2 R^2 = 1/R^2$. So $(x_1, \dots, x_m, y_0, y_1, \dots, y_t)$ has distance at most $\sqrt{5/8 + 1/R^2} < 1$ from the origin (assuming that $R \geq 2$), and we have found a common root of the f_i in $B^{m+1+t}(\mathbf{0}, 1)$. □

Corollary 3.10. $\mathsf{H_2N}$ *is* $\exists\mathbb{R}$*-complete.*

Proof. It is easy to see that $\mathsf{H_2N}$ belongs to $\exists\mathbb{R}$, so we only have to show that it is $\exists\mathbb{R}$-hard. By Lemma 3.9, deciding whether a family of polynomials $f_i : \mathbb{R}^n \to \mathbb{R}$, $i \in [s]$, has a common root $\mathbf{x} \in B^n(\mathbf{0}, 1)$ is $\exists\mathbb{R}$-complete, even if all f_i have total degree at most 2.

Using a standard transformation, we can turn each f_i into a homogeneous polynomial, just adding one additional variable y_0: Define

$$g_i(x_1, \dots, x_n, y_0) := y_0^4 f_i(x_1/y_0, \dots, x_n/y_0),$$

[15] This is a folklore result; for example, it is easy to see that STRETCHABILITY can be rephrased like this. The version in the Blum–Shub–Smale model can be found in [6].

$i \in [s]$. Then $(g_i)_{i \in [s]}$ is a family of homogeneous polynomials of total degree 4 and $(x_1, \ldots, x_n, y_0)$ is a common root of all g_i, $i \in [s]$, if and only if either $y_0 = 0$ or $y_0 \neq 0$ and $(x_1/y_0, \ldots, x_n/y_0)$ is a common root of the f_i, $i \in [s]$. This does not yet meet our goal, since we can have nontrivial common roots of the g_i that do not correspond to any common roots of the f_i: Just let $y_0 = 0$ and $(x_1, \ldots, x_n) \neq \mathbf{0}$. This can be remedied by making y_0 an upper bound of the x_i; let

$$g_{s+1}(x_0, x_1, \ldots, x_n, y_0) = y_0^4 - x_0^4 - \left(\sum_{i=1}^{n} x_i^2 \right)^2,$$

adding a new variable x_0.

Let $(x_0, x_1, \ldots, x_n, y_0)$ be a nontrivial common root of the g_i, $i \in [s+1]$ (adding x_0 to the list of variables of the g_i with $i \in [s]$). If $y_0 = 0$, then $x_0 = x_1 = \cdots = x_n = 0$ (because $g_{s+1} = 0$), so $(x_0, x_1, \ldots, x_n, y_0) = \mathbf{0}$ and the root is trivial. Hence, we must have $y_0 \neq 0$. But then $g_i(x_1, \ldots, x_n, y_0) = 0$ implies that $f_i(x_1/y_0, \ldots, x_n/y_0) = 0$, so $(x_1/y_0, \ldots, x_n/y_0)$ is a common root of the f_i. Moreover, since $g_{s+1} = 0$, we have that $y_0^4 > (\sum_{i=1}^n x_i^2)^2$, so $\sum_{i=1}^n (x_i/y_0)^2 < 1$, which implies that $(x_1/y_0, \ldots, x_n/y_0) \in B^n(\mathbf{0}, 1)$.

For the reverse direction, assume that we are given a root $(x_1, \ldots, x_n) \in B^n(\mathbf{0}, 1)$ of the f_i, $i \in [s]$. Let $y_0 := 1$ and $x_0 := (y_0^4 - (\sum_{i=1}^n x_i^2)^2)^{1/4}$ (which is defined, since $\sum_{i=1}^n x_i^2 < 1$). By definition, $g_{s+1}(x_0, x_1, \ldots, x_n, y_0) = 0$ and $g_i(x_0, x_1, \ldots, x_n, y_0) = 0$. So $(x_0, x_1, \ldots, x_n, y_0)$ is a nontrivial common root of the g_i, which is what we had to prove. □

Thus following Abbott's construction [1, Theorem 5.7],[16] we get $\forall\mathbb{R}$-hardness of rigidity.

Theorem 3.11. *Rigidity in $\mathbb{R}^2$ is $\forall\mathbb{R}$-complete.*

To complete the proof of Theorem 3.11, it remains to show that rigidity lies in $\forall\mathbb{R}$. The formal definition of rigidity given at the beginning of this section gives us an $\exists\forall$ formula for rigidity. We use a lemma that allows us to convert the leading existential quantifier into a group of universal quantifiers in this case.

Lemma 3.12. *Suppose*

$$\Phi(\varepsilon, y_1, \ldots, y_\ell) = (\forall x_1, \ldots, x_k)\ \varphi(\varepsilon, x_1, \ldots, x_k, y_1, \ldots, y_\ell),$$

is such that $\Phi(\varepsilon, y_1, \ldots, y_\ell)$ implies $\Phi(\varepsilon', y_1, \ldots, y_\ell)$ for all $\varepsilon > \varepsilon' > 0$. Then we can find a quantifier-free formula $\psi(\varepsilon, x_1, \ldots, x_k, y_1, \ldots, y_\ell, z_1, \ldots, z_m)$ of length at most $|\varphi| + dm$ in time $|\varphi| + dm$, where $m = cn^3 \log |\varphi|$, so that

$$(\exists \varepsilon > 0)(\forall x_1, \ldots, x_k)\ \varphi(\varepsilon, x_1, \ldots, x_k, y_1, \ldots, y_\ell)$$

[16] The reduction is polynomial time, since our polynomials have total degree 2.

is equivalent to

$$(\forall z_1,\ldots,z_m)(\forall \varepsilon)(\forall x_1,\ldots,x_k)\ \psi(\varepsilon,x_1,\ldots,x_k,y_1,\ldots,y_\ell,z_1,\ldots,z_m)$$

for some fixed constants $c,d>0$.

Remark 3.13. Lemma 3.12 is a consequence of the lemma by Grigor'ev and Vorobjov stated here as Theorem 1.3.

Proof of Theorem 3.11. We already showed $\forall\mathbb{R}$-hardness, based on the reduction by Abbott [1, Theorem 5.7]. We still have to show that the problem belongs to $\forall\mathbb{R}$. The input is a particular configuration of some linkage G given as a position $p(v)$ assigned to each vertex $V(G)=\{v_1,\ldots,v_n\}$. As a computational problem, the input needs to be finite, so we assume that the coordinates of $p(v)$ are algebraic real numbers. The configuration is rigid if there is an $\varepsilon>0$ so that for all $p'(v)$, $v\in V(G)$ with

(*i*) $|p'(v)-p(v)|\leq\varepsilon$ for all $v\in V$ and
(*ii*) $|p'(u)-p'(v)|=|p(u)-p(v)|$ for all $uv\in E(G)$,

we have that the configurations p' and p are congruent. This characterization is not $\forall\mathbb{R}$ because of the leading existential quantifier. However, it is monotone in ε in the sense of Lemma 3.12, so we can convert the existential quantifier into a block of universal quantifiers of length polynomial in n and the original formula. This if sufficient to show that the problem lies in $\forall\mathbb{R}$. □

Remark 3.14. The proof of Theorem 3.11 can probably be strengthened to show that recognizing the rigidity of unit distance graphs is $\forall\mathbb{R}$-complete. However, the construction will be a rather lengthy, technical recreation of Kempe's proof with unit distance graphs, so the author decided to leave this for a different occasion.

4 Open Questions

Approximation

Saxe introduced an approximate version of realizability for linkages, but we can apply it to both graph and linkage realizability: Call a graph or linkage ε-*approximately realizable* if the vertices can be placed so that for every edge, $1-\varepsilon<\text{len}(e)/\ell(e)<1+\varepsilon$, where $\text{len}(e)$ is the actual length of edge e in the straight-line drawing and $\ell(e)$ is the intended length. For unit distance graphs, this leads to the notion of a ε*-unit distance graph*.

Question 4.1. Is recognizing ε-approximately realizable graphs or linkages still $\exists\mathbb{R}$-hard? Is recognizing ε-unit distance graphs $\exists\mathbb{R}$-hard?

The case $\varepsilon = 1/2^{2^{O(n^C)}}$ would be a good starting point in view of Lemma 1.4. Saxe [40] showed that $1/9$-approximate realizability of linkages in $\mathbb{R}$ is **NP**-complete (this proof requires a different encoding from the one outlined in Remark 3.2, since the partition problem can be approximated well).

Higher Dimensions

All of our results were proved for $\mathbb{R}^2$; since $\mathbb{R}$ is well understood, this leaves the question of how hard it is to decide problems on linkages and graphs in higher-dimensional spaces.

Acknowledgements I would like to thank the anonymous referee for pointing out the universality papers by Jordan and Steiner [23] and Kapovich and Millson [25].

References

1. T.G. Abbott, Generalizations of Kempe's universality theorem. Master's thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, Cambridge, MA, 2008
2. S. Basu, R. Pollack, M.-F. Roy, Algorithms in real algebraic geometry, in *Algorithms and Computation in Mathematics*, vol. 10, 2nd edn. (Springer, Berlin, 2006)
3. D. Bienstock, Some provably hard crossing number problems. Discrete Comput. Geom. **6**(5), 443–459 (1991)
4. A. Björner, M.L. Vergnas, B. Sturmfels, N. White, G.M. Ziegler, Oriented matroids, in *Encyclopedia of Mathematics and Its Applications*, vol. 46, 2nd edn. (Cambridge University Press, Cambridge, 1999)
5. J. Blömer, A probabilistic zero-test for expressions involving roots of rational numbers, in *Algorithms—ESA '98 (Venice)*. Lecture Notes in Computer Science, vol. 1461 (Springer-Verlag, Berlin, 1998), pp. 151–162
6. L. Blum, M. Shub, S. Smale, On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines. Bull. Am. Math. Soc. (N.S.) **21**(1), 1–46 (1989)
7. S. Cabello, E.D. Demaine, G. Rote, Planar embeddings of graphs with specified edge lengths. J. Graph Algorithm. Appl. **11**(1), 259–276 (electronic) (2007)
8. J. Canny, Some algebraic and geometric computations in pspace, in *STOC '88: Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing* (ACM, New York, 1988), pp. 460–469
9. V. Chvátal, D.A. Klarner, D.E. Knuth, Selected combinatorial research problems. Technical Report STAN-CS-72-292, Computer Science Department, Stanford University, Palo Alto, CA, 1972
10. R. Connelly, E.D. Demaine, Geometry and topology of polygonal linkages, in *Handbook of Discrete and Computational Geometry*, ed. by J.E. Goodman, J. O'Rourke (CRC Press, West Palm Beach, FL, 2004)
11. E.D. Demaine, J. O'Rourke, *Geometric Folding Algorithms* (Cambridge University Press, Cambridge, 2007); Linkages, origami, polyhedra

12. G. Di Battista, L. Vismara, Angles of planar triangular graphs. SIAM J. Discrete Math. **9**(3), 349–359 (1996)
13. P. Eades, N.C. Wormald, Fixed edge-length graph drawing is NP-hard. Discrete Appl. Math. **28**(2), 111–134 (1990)
14. D. Eppstein, Unit distance graphs (2010), Blog posting at http://11011110.livejournal.com/188807.html. Accessed 10 September 2010
15. P. Erdős, M. Simonovits, On the chromatic number of geometric graphs. Ars Combin. **9**, 229–246 (1980)
16. P. Erdős, F. Harary, W.T. Tutte, On the dimension of a graph. Mathematika **12**, 118–122 (1965)
17. X. Gao, C. Zhu, Automated generation of Kempe linkage and its complexity. J. Comput. Sci. Tech. **14**(5), 460–467 (1999)
18. J.E. Goodman, J. O'Rourke (eds.), *Handbook of Discrete and Computational Geometry*, 2nd edn. Discrete Mathematics and Its Applications (Chapman & Hall/CRC, Boca Raton, FL, 2004)
19. J.E. Goodman, R. Pollack, B. Sturmfels, Coordinate representation of order types requires exponential storage, in *STOC '89: Proceedings of the Twenty-First Annual ACM Symposium on Theory of Computing* (ACM, New York, 1989), pp. 405–410
20. D.Yu. Grigor'ev, N.N. Vorobjov, Solving systems of polynomial inequalities in subexponential time. J. Symb. Comput. **5**(1–2), 37–64 (1988)
21. S.L. Greitzer, H.S.M. Coxeter, *Geometry Revisited* (The Mathematical Association of America, Washington, DC, 1967)
22. H. Hong. Comparison of several decision algorithms for the existential theory of the reals. Technical Report 91–41, RISC-Linz, Johannes Kepler University, Linz, 1991
23. D. Jordan, M. Steiner, Configuration spaces of mechanical linkages. Discrete Comput. Geom. **22**(2), 297–315 (1999)
24. M. Kahle, Algorithm for embedding a graph with metric constraints, 2010. Posting on mathoverflow at http://mathoverflow.net/questions/33043. Accessed 9 Sept 2010
25. M. Kapovich, J.J. Millson, Universality theorems for configuration spaces of planar linkages. Topology **41**(6), 1051–1107 (2002)
26. A.B. Kempe, How to draw a straight line. *Classics in Mathematics Education* (MacMillan, London, 1977)
27. P. Koiran, The complexity of local dimensions for constructible sets. J. Complexity **16**(1), 311–323 (2000); Real computation and complexity (Schloss Dagstuhl, 1998)
28. M. Laurent, A tour d'horizon on positive semidefinite and Euclidean distance matrix completion problems. Topics in semidefinite and interior-point methods (Toronto, ON, 1996), in *Fields Institute of Communication*, vol. 18 (Amer. Math. Soc., Providence, RI, 1998), pp. 51–76
29. P. Lemke, S.S. Skiena, W.D. Smith, Reconstructing sets from interpoint distances. Discrete and computational geometry, in *Algorithms Combination*, vol. 25 (Springer, Berlin, 2003), pp. 507–631
30. D. Little, Peaucellier's linkage (2005), Applet at 9 october 2012 http://www.math.psu.edu/dlittle/java/parametricequations/linkage.html. Accessed 21 Sept 2012
31. H. Maehara, Note on induced subgraphs of the unit distance graph E^n. Discrete Comput. Geom. **4**(1), 15–18 (1989)
32. C.J.H. McDiarmid, T. Müller, The number of bits needed to represent a unit disk graph. Graph-theoretic concepts in computer science. Lecture Notes in Comput. Sci. vol. 6410 (Springer, Berlin, 2010), pp. 315–323 http://dx.doi.org/10.1007/978-3-642-16926-7_29
33. B. Mishra, Computational real algebraic geometry, in *Handbook of Discrete and Computational Geometry*, CRC Press Ser. Discrete Math. Appl. (CRC Press, Boca Raton, FL, 1997), pp. 537–556
34. N. E. Mnëv, The universality theorems on the classification problem of configuration varieties and convex polytopes varieties. Topology and Geometry—Rohlin Seminar, in *Lecture Notes in Mathematics*, vol. 1346 (Springer-Verlag, Berlin, 1988), pp. 527–543
35. C.H. Papadimitriou, *Computational Complexity* (Addison-Wesley, Reading, MA, 1994)

36. J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. I. Introduction. Preliminaries. The geometry of semi-algebraic sets. The decision problem for the existential theory of the reals. J. Symbolic Comput. **13**(3), 255–299 (1992)
37. J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. II. The general decision problem. Preliminaries for quantifier elimination. J. Symbolic Comput. **13**(3), 301–327 (1992)
38. J. Renegar, On the computational complexity and geometry of the first-order theory of the reals. III. Quantifier elimination. J. Symbolic Comput. **13**(3), 329–352 (1992)
39. J. Richter-Gebert, Mnëv's universality theorem revisited. Sém. Lothar. Combin. **34** (1995)
40. J.B. Saxe, Embeddability of weighted graphs in *k*-space is strongly NP-hard, in *Seventeenth Annual Allerton Conference on Communication, Control, and Computing, Proceedings of the Conference held in Monticello, Ill.*, Urbana, 1979. University of Illinois Department of Electrical Engineering. Proceedings of the International School of Physics "Enrico Fermi", LXX*, pp. xiv+1036, 10–12 October 1979
41. M. Schaefer, Complexity of some geometric and topological problems, in *Graph Drawing*, ed. by D. Eppstein, E.R. Gansner. Lecture Notes in Computer Science, vol. 5849 (Springer, Berlin, 2009), pp. 334–344
42. M. Schaefer, D. Štefankovič, Fixed points, Nash equilibria, and the existential theory of the reals. Unpublished manuscript.
43. P.W. Shor, Stretchability of pseudolines is NP-hard. Applied geometry and discrete mathematics, in *DIMACS Series in Discrete Mathematical and Theoretical Computer Science*, vol. 4 (Amer. Math. Soc., Providence, RI, 1991), pp. 531–554
44. M. Sipser, *Introduction to the Theory of Computation*, 2nd edn. (Course Technology, Boston, 2005)
45. A. Soifer, *The Mathematical Coloring Book* (Springer, New York, 2009); Mathematics of coloring and the colorful life of its creators, With forewords by Branko Grünbaum, Peter D. Johnson, Jr. and Cecil Rousseau
46. J.J. Sylvester, On recent discoveries in mechanical conversion of motion, in *Proceedings of the Royal Institution of Great Britain*, London, 1875, pp. 179–198; Talk delivered 23 January 1874
47. L.A. Székely, A successful concept for measuring non-planarity of graphs: the crossing number. Discrete Math. **276**(1–3), 331–352 (2004)
48. S. Whitesides, Algorithmic issues in the geometry of planar linkage movement. Australian Comput. J. **24**(2), 42–50 (1992)
49. Y. Yemini, Some theoretical aspects of position-location problems, in *20th Annual Symposium on Foundations of Computer Science (San Juan, Puerto Rico, 1979)* (IEEE, New York, 1979), pp. 1–8

Equilateral Sets in ℓ_p^d

Clifford Smyth

Abstract If X is a Minkowski space, i.e., a finite-dimensional real normed space, then $S \subset X$ is an equilateral set if all pairs of points of S determine the same distance with respect to the norm. Kusner conjectured that $e(\ell_p^d) = d+1$ for $1 < p < \infty$ and $e(\ell_1^d) = 2d$ [6]. Using a technique combining linear algebra and approximation theory, we prove that for all $1 < p < \infty$, there exists a constant $C_p > 0$ such that $e(\ell_p^d) \leq C_p d^{1+2/(p-1)}$.

Mathematics Subject Classification (2010): 52C17(41A10, 46B20, 52A21)

1 Introduction

Let $(X, \|\cdot\|)$ be a finite-dimensional real normed space, also called a *Minkowski space*. Let $\Delta(S) := \{\|s-t\| : s,t \in S, s \neq t\}$. For $k \geq 1$, we say S is a k-distance set if $|\Delta(S)| \leq k$. Let $e_k(X)$ be the maximum possible cardinality of a k-distance set in X. This maximum exists: $e_k(X) \leq (k+1)^{\dim(X)}$ [11]. We call a 1-distance set an *equilateral* set and write $e(X) := e_1(X)$.

Petty showed that $e(X) \leq 2^{\dim(X)}$ and that equality is obtained if and only if X is affinely isometric to ℓ_∞^d [10]. Brass showed that $e(X) = \Omega((\log d/\log\log d)^{1/3})$ [4]. Subsequently, Swanepoel and Villa showed that $e(X) \geq \exp(c\sqrt{d})$ for some $c > 0$ [13]. It is an open question whether $e(X) \geq \dim(X)+1$ [5]. This has been proved for $\dim(X) = 3,4$ [8, 10].

C. Smyth (✉)
Department of Mathematics and Statistics, 116 Petty Building, University of North Carolina Greensboro, Greensboro, NC 27402, USA
e-mail: cdsmyth@uncg.edu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_25,

We will consider the case where $X = \ell_p^d$, i.e., $\mathbb{R}^d$ in the L^p-norm. Kusner conjectured that $e(\ell_p^d) = d+1$ for $1 < p < \infty$ and $e(\ell_1^d) = 2d$ [6]. It is well known that $e(\ell_2^d) = d+1$. We have $e(\ell_p^d) \geq d+1$ as the standard basis vectors together with an appropriate scalar multiple of $(1,1,\ldots,1)$ form an equilateral set in ℓ_p^d. Using a technique combining linear algebra and approximation theory, we show that

Theorem 1. *For all $1 < p < \infty$, there is a constant $C_p > 0$ such that*

$$e(\ell_p^d) \leq C_p d^{1+2/(p-1)}.$$

This was later improved by Alon and Pudlák to $e(\ell_p^d) \leq C'_p d^{1+3/(2p-1)}$ for $1 \leq p < \infty$ and $e(\ell_p^d) \leq C''_p d \log d$ for $p \geq 1$, an odd integer [1]. Here C'_p and C''_p are positive constants. Galvin noted that $e(\ell_p^d) \leq 1+(p-1)d$ for p an even integer (personal communication, 1999). Swanepoel proved $e(\ell_p^d) \leq (2\lceil p/4 \rceil - 1)d+1$ for p an even integer [12]. In particular, $e(l_4^d) = d+1$. Swanepoel also proved (in particular) that for every $1 \leq p < 2$, $e(\ell_p^d) > d+1$ for d large enough [12].

We will prove Theorem 1 and then make some conjectures concerning $e_k(X)$.

2 Proof of Theorem 1

Before proving Theorem 1, we require the following three easy or well-known results. Let Mat_m be the set of $m \times m$ real matrices. For $M \in \mathrm{Mat}_m$, let $\|M\|_\infty := \max_{1 \leq i,j \leq m} |M_{ij}|$. Let I_m denote the $m \times m$ identity matrix.

Lemma 2.1 ([2]). *Let V be the vector space of real-valued functions on a set X. Let $\{f_1,\ldots,f_m\} \subset V$. Let $\{a^1,\ldots,a^m\} \subset X$. Let $M \in Mat_m$ be the matrix with $M_{ij} = f_i(a^j)$. If M is invertible, then the f_i are linearly independent.*

Lemma 2.2 ([14]). *If $M \in Mat_m$ and $\|M - I_m\|_\infty < 1/m$, then M is invertible.*

Lemma 2.3. *Let $1 < p < \infty$. There is a sequence of polynomials $\{q_l(x)\}_{l \geq \lceil p \rceil}$ with* $\deg(q_l) \leq l$, *such that $||x|^p - q_l(x)| \leq B_p/l^p$, for all $|x| \leq 1$ where $B_p > 0$ is a constant.*

This last lemma is an application of the following result of Jackson [7]. For the statement given here, see p. 57 of [9].

Theorem 2. *Let $k \geq 1$. Suppose $f \in C^k([-1,1])$ and $f^{(k)} \in Lip_M\, \alpha$. For every $l > k$, there is a polynomial q_l of degree at most l such that $\|f - q_l\|_\infty \leq D(k,\alpha,l)c^{k+1}M/l^{k+\alpha}$, where $D(k,\alpha,l) = l^{k+\alpha}/((l)_k(l-k)^\alpha)$ and $c = 1+\pi^2/2$.*

Here $\mathrm{Lip}_M\, \alpha$ is the class of functions $f(x)$ on $[-1,1]$ such that $|f(x)-f(y)| \leq M|x-y|^\alpha$ for all $x,y \in [-1,1]$ and $\|f(x)\|_\infty = \max_{x\in[-1,1]} f(x)$. As is common notation, $(l)_k$ is the falling factorial, $(l)_k := l(l-1)\cdots(l-k+1)$.

To obtain Lemma 2.3, we first note that each factor of $D(k,\alpha,l) = \prod_{i=1}^{k-1} (1+i/(l-i))(1+k/(l-k))^\alpha$ is decreasing with l. We thus obtain an upper bound for this quantity by setting $l = k+1$, namely, $D(k,\alpha,l) \le (k+1)^{k+\alpha}/(k+1)!$ If $f(x) = |x|^p$, set $k = \lceil p \rceil - 1$ and $\alpha = p - k \in (0,1]$. Then $f \in C^k([-1,1])$ and $f^{(k)}(x) = \operatorname{sgn}^k(x)(p)_k|x|^\alpha \in \operatorname{Lip}_M \alpha$, where $M = (p)_k$. Thus, we obtain Lemma 2.3 with $B_p = (\lceil p \rceil^p (1+\pi^2/2)^{\lceil p \rceil}(p)_{\lceil p \rceil -1})/\lceil p \rceil !$. It is straightforward to verify that $B_p \ge \lceil p \rceil^p$, a fact that we require in the proof of Theorem 1.

Proof (Theorem 1.). Fix $p \in (1,\infty)$. Let $S = \{a^1, a^2, \ldots, a^m\} \subset \ell_p^d$ be an equilateral set of maximum size, scaled so that $\|a^i - a^j\|_p = 1$ for all $i \ne j$. We define the following functions from $\mathbb{R}^n$ to $\mathbb{R}$, $f_i(x) = 1 - \|x - a^i\|_p^p = 1 - \sum_{t=1}^d |x_t - a_t^i|^p$, $i = 1,\ldots,m$ and $g_i(x) = 1 - \sum_{t=1}^d q_l(x_t - a_t^i)$, $i = 1,\ldots,m$, where q_l is the polynomial from Lemma 2.3 approximating $|x|^p$ to within B_p/l^p on $[-1,1]$ and where l is smallest integer such that $\mathrm{d}B_p/l^p < 1/m$. Note that $l > (\mathrm{d}B_p m)^{1/p} \ge B_p^{1/p} \ge \lceil p \rceil$, as required by Lemma 2.3. Since $B_p \ge 1$, $(\mathrm{d}B_p m)^{1/p} \ge 1$ and $l \le (\mathrm{d}B_p m)^{1/p} + 1 \le 2(\mathrm{d}B_p m)^{1/p}$.

Let $M = [g_i(a^j)] \in \mathrm{Mat}_m$. By the assumption on S, we have $[f_i(a^j)] = I_m$. Let $A = M - I_m = [g_i(a^j) - f_i(a^j)]$. We have $\left|A_{ij}\right| \le \sum_{t=1}^d |\zeta_{ijt}|$, where $\zeta_{ijt} := |a_t^j - a_t^i|^p - q_l(a_t^j - a_t^i)$. Since $|a_t^j - a_t^i| \le \|a^j - a^i\|_p \le 1$, we have $\left|\zeta_{ijt}\right| \le B_p/l^p$ by Lemma 2.3. Thus, $\|M - I_m\|_\infty \le \mathrm{d}B_p/l^p < 1/m$. This implies that M is invertible by Lemma 2.2, and thus the g_i are linearly independent by Lemma 2.1.

The g_i are elements of W, the subspace of $\mathbb{R}[x_1,\ldots,x_d]$ spanned by $\{1, \sum_{t=1}^d x_t^l\} \cup \bigcup_{t=1}^d \{x_t, x_t^2, \ldots, x_t^{l-1}\}$. Thus, we have $m \le \dim(W) = 2 + d(l-1) \le dl$, if $d \ge 2$. Since $l \le 2(\mathrm{d}B_p m)^{1/p}$, $m \le 2d(\mathrm{d}B_p m)^{1/p}$, or $e(\ell_p^d) = m < 2^{p/(p-1)} B_p^{1/(p-1)} d^{(p+1)/(p-1)}$. □

Notes: (1) The constant $2^{p/(p-1)} B_p^{1/(p-1)}$ grows without bound as p approaches 1 and is $O(p)$ as p goes to infinity. (2) As Galvin noted (personal communication, 1999), if p is an even integer, the functions $f_i(x)$ *are* polynomials of degree p and, together with the identity function, 1, form a linearly independent set in W. Thus, $e(\ell_p^d) \le 1 + d(p-1)$.

3 *k*-Distance Sets

For $d \ge 1$, $e(X)$ is upper semicontinuous on the Banach–Mazur compactum of normed spaces of dimension d. For a proof, suppose N_n is a sequence of norms converging to the norm N in the Banach–Mazur distance and $S_n = \{a_{n,1}, \ldots, a_{n,e}\}$ is an equilateral set of size e in N_n with $\Delta(S_n) = \{1\}$. Then there is an equilateral set S of size e in N. Indeed, we may assume by translating each S_n that $\cup S_n$ lies in a compact set. By passing to convergent subsequences, we get $a_{n,i} \to a_i$ for all $1 \le i \le e$. $S = \{a_1, \ldots, a_n\}$ is equilateral in N. A particular instance is $e(X) \le d+1$ for X sufficiently close to ℓ_2^d.

It seems plausible that

Conjecture 1. For all $d,k \geq 1$, $e_k(X)$ is upper semicontinuous.

Trying the same approach used for $e(X)$ leads to a problem. Setting $\Delta(S_n) = \{1 = d_{n,1} > d_{n,2} > \cdots > d_{n,k}\}$, we can pass to subsequences to ensure $\|a_{n,i} - a_{n,j}\|_n = d_{f(i,j)}$, where f is a fixed function, $d_{n,i} \to d_i$, and $\|a_i - a_j\| = d_{f(i,j)}$, but we may have $d_k = 0$ and $|S| < e$. This state of affairs would be rectified by proving the following conjecture.

Conjecture 2. For all $d,k \geq 1$, there is a universal constant $c_{d,k} > 0$ so that for any normed space X of dimension d, there is a maximum size k-distance set $S \subset X$ with maximum distance 1 and minimum distance greater than or equal to $c_{d,k}$.

We'd like to prove a k-distance analogue of Theorem 1. It is known that $\binom{d+1}{k} \leq e_k(\ell_2^d) \leq \binom{d+k}{k}$ [3]. Mimicking the proof of the lower bound in the previous statement, it is trivial to show that the $\binom{d}{k}$ 0–1 vectors of length d with exactly k 1's form a k-distance set in ℓ_p^d, so that $e_k(\ell_p^d) \geq \binom{d}{k}$.

Conjecture 3. For all $k \geq 1$, there exists $p(k)$ so that for all $p > p(k)$, there exists a constant $C_{p,k} > 0$ so that $e_k(\ell_p^d) \leq C_{p,k} d^k$.

The natural attempt at a proof of this conjecture would be to define for $S = \{a^1, \ldots, a^m\} \subset \ell_p^d$ the functions $f_i(x) = \prod_{s=1}^k \left(d_s^p - \sum_{t=1}^d |x_t - a_t^i|^p\right)$ and the polynomial approximations $g_i(x) = \prod_{s=1}^k \left(d_s^p - \sum_{t=1}^d q_l(x_t - a_t^i)\right)$, where $\Delta(S) = \{1 = d_1 > d_2 > \cdots > d_k\}$. However, when trying to prove $M = [g_i(a^j)]$ is invertible, the sufficient condition $\|M - [f_i(a^j)]\|_\infty = \|M - d_1^p \ldots d_k^p I_m\|_\infty < d_1^p \ldots d_k^p / m$ involves an estimate that could require an arbitrarily high degree of approximation l, as d_k could be arbitrarily small. One needs lower bounds on d_k in order to make this approach work.

The following is a seemingly reasonable conjecture.

Conjecture 4. For all $p \geq 1$ and all $k \geq 1$, there exists a constant $c_{p,k} > 0$ independent of d so that there exists a k-distance set $S \subset \ell_p^d$ with $|S| = e_k(\ell_p^d)$ and with $\Delta(S) = \{1 = d_1 > d_2 > \cdots > d_k \geq c_{p,k}\}$.

Assuming this conjecture, we can prove

Conjecture 5. For all $k \geq 1$ and for all $p > k$, there exists a constant $C_{p,k} > 0$ so that $e_k(l_p^d) \leq C_{p,k} d^{k+((k^2+k)/(p-k))}$.

Acknowledgements The author would like to thank Konrad Swanepoel, Michael Saks, David Galvin, and Noga Alon for very helpful conversations.

References

1. N. Alon, P. Pudlák, Equilateral sets in ℓ_p^n. Geomet. Funct. Anal. **13**(3), 467–482 (2003)
2. L. Babai, P. Frankl, Linear algebra methods in combinatorics. Manuscript (1992). http://www.cs.uchicago.edu/research/publications/combinatorics

3. A. Blokhuis, *Few Distance Sets*. C.W.I Tracts no. 7 (Mathematisch Centrum, Amsterdam, 1984)
4. P. Brass, On equilateral simplices in normed spaces. Cont. Alg. Geom. **40**, 303–307 (1999)
5. B. Grünbaum, On a conjecture of H. Hadwiger. Pacific J. Math. **11**, 215–219 (1961)
6. R. Guy (ed.), Unsolved problems: an olla-podrida of open problems, often oddly posed. Am. Math. Mon. **90**, 196–200 (1983)
7. D. Jackson, *The Theory of Approximation* (American Mathematical Society, New York, 1930)
8. V.V. Makeev, Equilateral simplices in a four-dimensional normed space. J. Math. Sci. (N.Y.) **140**, 548–550 (2007)
9. G. Meinardus, *Approximation of Functions: Theory and Numerical Methods* (Springer-Verlag, New York, 1967)
10. C.M. Petty, Equilateral sets in Minkowski spaces. Proc. AMS **29**, 369–374 (1971)
11. K.J. Swanepoel, Cardinalities of k-distance sets in Minkowski spaces. Discrete Math. **197/198**, 759–767 (1999)
12. K.J. Swanepoel, A problem of Kusner on equilateral sets. Arch. Math. **83**, 164–170 (2004)
13. K.J Swanepoel, R. Villa. A lower bound for the equilateral number of normed spaces. Proc. Am. Math. Soc. **136**, 127–131 (2008)
14. O. Taussky, A recurring theorem on determinants. Am. Math. Mon. **56**(10), 672–676 (1949)

A Note on Geometric 3-Hypergraphs

Andrew Suk

Abstract In this note, we prove several Turán-type results on geometric hypergraphs. The two main theorems are (1) every n-vertex geometric 3-hypergraph in the plane with no three strongly crossing edges has at most $O(n^2)$ edges, and (2) every n-vertex geometric 3-hypergraph in 3-space with no two disjoint edges has at most $O(n^2)$ edges. These results support two conjectures that were raised by Dey and Pach, and by Akiyama and Alon.

1 Introduction

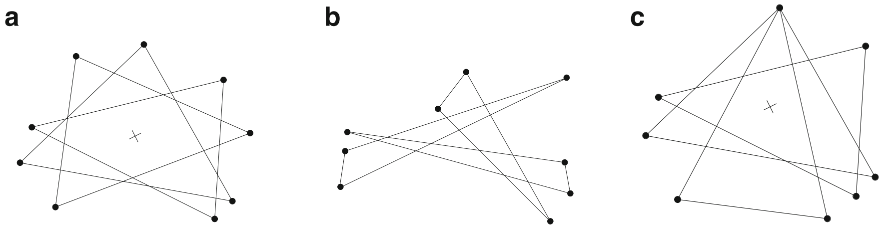
A *geometric r-hypergraph H in d-space* is a pair (V,E), where V is a set of points in general position in Euclidean d-space, and E is a set of closed $(r-1)$-dimensional simplices (edges) induced by some r-tuple of V. The sets V and E are called the *vertex set* and *edge set* of H, respectively. Two edges in H are *crossing* if they are vertex disjoint and have a point in common. Notice that if k edges are pairwise crossing, it does not imply that they all have a point in common. Hence, we say that H contains k *strongly crossing edges* if H contains k vertex disjoint edges that all share a point in common. See Fig. 1.

A direct application of the colored Tverberg theorem (see [3, 19]) gives

Theorem 1. *Let* $\mathrm{ex}_d(SC_k^{d+1},n)$ *denote the maximum number of edges an n-vertex geometric* $(d+1)$*-hypergraph in d-space has with no k strongly crossing edges. Then*

A. Suk (✉)
Massachusetts Institute of Technology, Department of Mathematics, 77 Massachusetts Ave., 4-307 Cambridge, MA, 02139, USA
e-mail: asuk@math.mit.edu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_26,



a. Three strongly crossing edges.

b. Three pairwise crossing edges with an empty intersection.

c. Three edges not strongly crossing since two share a vertex.

Fig. 1 Three edges of a geometric 3-hypergraph in the plane

$$\mathrm{ex}_d(SC_k^{d+1},n) = O\left(n^{d+1-\frac{1}{(2k-1)^d}}\right).$$

Dey and Pach [5] showed that $\mathrm{ex}_d(SC_2^{d+1},n) = \Theta(n^d)$, and conjectured $\mathrm{ex}_d(SC_k^{d+1},n) = \Theta(n^d)$ for every fixed d and k. The lower bound can easily be seen by taking all edges with a vertex in common. The main motivation for their conjecture is in deriving upper bounds on the maximum number of *k-sets* of an n-point set in $\mathbb{R}^d$. See [11] for more details. In this note, we settle the Dey–Pach conjecture for geometric 3-hypergraphs in the plane with no three strongly crossing edges, and improve the upper bound of $\mathrm{ex}_2(SC_k^3,n)$.

Theorem 2. $\mathrm{ex}_2(SC_3^3,n) = \Theta(n^2)$.

Theorem 3. *For fixed* $k \geq 4$, $\mathrm{ex}_2(SC_k^3,n) \leq O(n^{3-\frac{1}{k}})$.

As a related result, Akiyama and Alon [2] used the Borsuk–Ullam theorem [4] to show the following.

Theorem 4. *Let* $\mathrm{ex}_d(D_k^d,n)$ *denote the maximum edges that an n-vertex geometric d-hypergraph in d-space has with no k pairwise disjoint edges. Then*

$$\mathrm{ex}_d(D_k^d,n) \leq n^{d-(1/k)^{d-1}}.$$

They conjecture that for every fixed d and k, $\mathrm{ex}_d(D_k^d,n) = \Theta(n^{d-1})$. Again, the lower bound can easily be seen by taking all edges with a vertex in common. Pach and Törőcsik [14] showed that $\mathrm{ex}_2(D_k^2,n) = O(k^4 n)$, which was later improved to $O(k^2 n)$ by Tóth [16]. Here we settle the Akiyama–Alon conjecture for geometric 3-hypergraphs in 3-space with no two disjoint edges.

Theorem 5. $\mathrm{ex}_3(D_2^3,n) = \Theta(n^2)$.

For clarity of the proofs, we do not make any attempts to optimize the constants.

2 Strongly Crossing Edges in the Plane

In this section, we will prove Theorems 2 and 3. Recall that a *geometric graph* is a graph drawn in the plane with vertices represented by points and edges by straight-line segments connecting the corresponding pairs. Recently, Ackerman [1] showed the following.

Lemma 2.1. *Let $G = (V,E)$ be an n-vertex geometric graph in the plane with no four pairwise crossing edges. Then $|E(G)| \leq O(n)$.* □

We note that Lemma 2.1 holds for topological graphs. Before we give the proofs, we will introduce some terminology. Consider a family $\mathcal{S} = \{s_1,\dots,s_k\}$ of pairwise crossing segments in the plane, and let $\mathcal{L} = \{l_1,\dots,l_k\}$ be a family of lines such that l_i is the line supported by segment s_i. Recall that the *level* of a point $x \in \cup\mathcal{L}$ is defined as the number of lines of $\mathcal{L}$ lying strictly below x. We define the *top level of* $\mathcal{L}$ as the closure of the set of points in $\cup\mathcal{L}$ with level $k-1$. We define the *top level of* $\mathcal{S}$ to be the top level of $\mathcal{L}$. See Fig. 2. Notice that L is a (not strictly) convex function.

For each edge t in a geometric 3-hypergraph in the plane, we define its *base* as the side with the longest x-projection. We define the other two sides of t as its *left* and *right* sides. See Fig. 3. Notice that every edge in a geometric 3-hypergraph is incident to a vertex that lies strictly above or below its base. We are now ready to prove Theorem 2.

Proof of Theorem 1.2. Let $H = (V,E)$ be an n-vertex geometric 3-hypergraph in the plane with no three strongly crossing edges. We can assume that $|E(H)| \geq 20n^2$ (since otherwise we would be done) and at most $|E(H)|/2$ edges in H are incident to a vertex that lies strictly below its base. We will discard all such edges, leaving us with at least $|E(H)|/2$ edges left. Let E_{uv} be the set of edges in H with base uv.

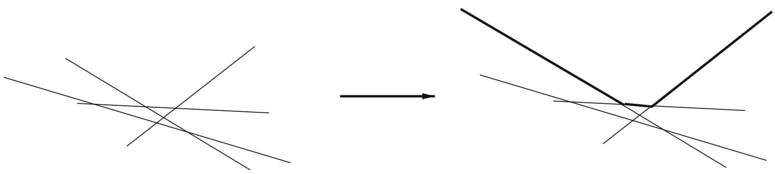


Fig. 2 The top level of four pairwise crossing segments is drawn *thick*

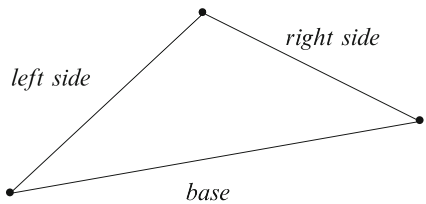


Fig. 3 The base, left side, and right side

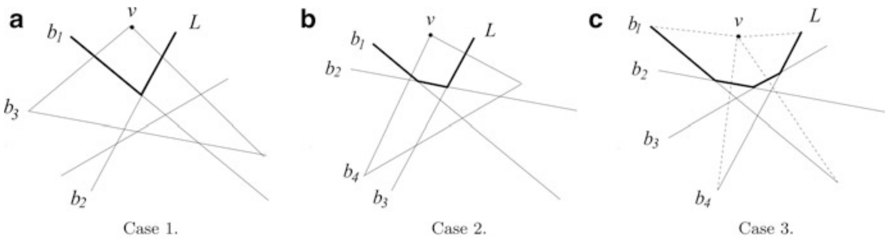


Fig. 4 Three cases

We discard all sets E_{uv} for which $|E_{uv}| \leq |E(H)|/(2n^2)$. Since we have thrown away at most $|E(H)|/4$ edges in this process, we have at least $|E(H)|/4$ edges left. Therefore, $|E_{uv}| = 0$ or $|E_{uv}| \geq |E(H)|/(2n^2) \geq 10$.

Now let $G_v = (V,E)$ denote the geometric graph with $V(G_v) = V(H)$ and $xy \in E(G_v)$ if $\mathrm{conv}(x \cup y \cup v) \in E(H)$ with base xy.

Observation 2.2. *G_v does not contain four pairwise crossing edges (bases).*

Proof. For sake of contradiction, suppose G_v contains four pairwise crossing edges $b_1, b_2, b_3, b_4 \in E(G_v)$. Then v lies above b_i for all i. Let L denote the top level of the arrangement $\mathcal{S} = \{b_1, b_2, b_3, b_4\}$. Now the proof falls into three cases.

Case 1. Suppose L intersects exactly two members of $\mathcal{S}$, say, bases b_1 and b_2 (in order from left to right along L). Let p be the intersection point of b_1 and b_2. Then the vertical line through p must intersect b_3 below p. Moreover, since segments b_1 and b_3 cross, v and the right endpoint of b_3 must lie on the same half-plane generated by the line supported by b_1. Likewise, v and the left endpoint of b_3 must lie on the same half-plane generated by the line supported by b_2. Therefore, $p \in \mathrm{conv}(v \cup b_3)$. See Fig. 4a. Since $|E_{b_1}|, |E_{b_2}| \geq 10$, there exist vertices $x, y \in V(H)$ such that $\mathrm{conv}(v \cup b_3), \mathrm{conv}(x \cup b_1), \mathrm{conv}(y \cup b_2)$ are three (vertex disjoint) strongly crossing edges in H, and we have a contradiction.

Case 2. Suppose L intersects exactly three members of $\mathcal{S}$, say, bases b_1, b_2, b_3 (in order from left to right along L). Now b_4 must intersect b_2 to either the left or right of $b_2 \cap L$. Without loss of generality, we can assume that b_4 intersects b_2 to the right of $b_2 \cap L$. Let p be the intersection point of segments b_2 and b_3. By the same argument as above, $p \in \mathrm{conv}(v \cup b_4)$. See Fig. 4b. Since $|E_{b_1}|, |E_{b_2}| \geq 10$, there exist vertices $x, y \in V(H)$ such that $\mathrm{conv}(v \cup b_4), \mathrm{conv}(x \cup b_1), \mathrm{conv}(y \cup b_2)$ are three strongly crossing edges in H, and we have a contradiction.

Case 3. Suppose L intersects b_1, b_2, b_3, b_4 in order from left to right along L. Let p be the intersection point of segments b_2 and b_3, and let l be the vertical line through v. Since the right endpoint of b_4 lies to the right of l, and the left endpoint of b_1 lies to the left of l, we have $p \in \mathrm{conv}(v \cup b_1) \cup \mathrm{conv}(v \cup b_4)$. Therefore, either $\mathrm{conv}(v \cup b_1)$

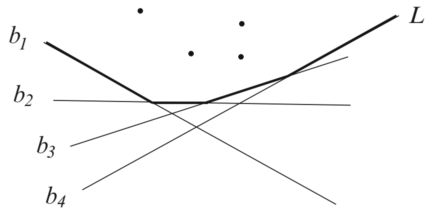


Fig. 5 Arrangement of b_1, b_2, b_3, b_4

or $\text{conv}(v \cup b_4)$ (say $\text{conv}(v \cup b_1)$) contains p. See Fig. 4c. Since $|E_{b_2}|, |E_{b_3}| \geq 10$, there exist vertices $x, y \in V(H)$ such that $\text{conv}(v \cup b_1), \text{conv}(x \cup b_2), \text{conv}(y \cup b_3)$ are three strongly crossing edges in H, and we have a contradiction. □

Therefore, by Lemma 2.1, $|E(G_v)| \leq O(n)$ for every vertex $v \in V(H)$. Hence,

$$\frac{|E(H)|}{4} \leq \sum_{v \in V(H)} |E(G_v)| = O(n^2),$$

which implies $|E(H)| = O(n^2)$. □

Before we prove Theorem 3, we will need the following lemma due to Valtr [17].

Lemma 2.3. *Let $G = (V, E)$ be an n-vertex geometric graph in the plane such that all of the edges in G intersect the y-axis. If G does not contain k pairwise crossing edges, then $|E(G)| \leq c_k n$, where c_k depends only on k.* □

Proof of Theorem 1.3. Let H be an n-vertex geometric 3-hypergraph in the plane with no k strongly crossing edges for $k \geq 4$. Just as before, we can assume at most $|E(H)|/2$ of the edges in H are incident to a vertex that lies strictly below its base. We discard all such edges, leaving us with at least $|E(H)|/2$ edges left in H. Now we make the following observation.

Observation 2.4. *Suppose $b_1, \ldots, b_k$ are k pairwise crossing bases and $v_1, \ldots, v_k \in V(H)$ such that $\text{conv}(v_i \cup b_j) \in E(H)$ with base b_j for all i, j. Then H contains k strongly crossing edges.*

Proof. Let L denote the top level of the segment arrangement $\mathcal{S} = \{b_1, \ldots, b_k\}$, and assume that $b_1, \ldots, b_k$ are ordered by increasing slopes. See Fig. 5.

Now we define edges $t_1, t_2, \ldots, t_k \in E(H)$ as follows. Among the k edges $\text{conv}(b_1 \cup v_1), \text{conv}(b_1 \cup v_2), \ldots, \text{conv}(b_1 \cup v_k) \in E(H)$ (with slight abuse of notation), let $t_1 = \text{conv}(b_1 \cup v_1)$ be the edge whose right side has the rightmost intersection with L. Then among the $k-1$ edges $\text{conv}(b_2 \cup v_2), \text{conv}(b_2 \cup v_3), \ldots, \text{conv}(b_2 \cup v_k)$ (again with slight abuse of notation), let $t_2 = \text{conv}(b_2 \cup v_2)$ be the edge whose right side has the rightmost intersection with L. We continue this procedure until we have k edges $t_1, t_2, \ldots, t_k$. Clearly, these k edges are vertex disjoint.

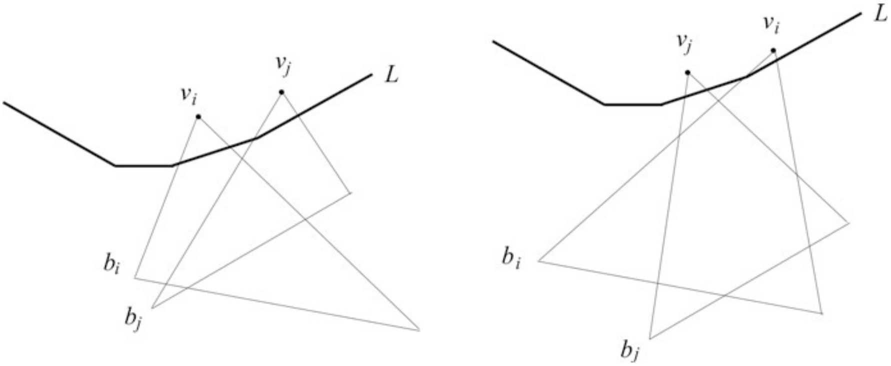


Fig. 6 Assume $(t_i \cap L) \cap (t_j \cap L) = \emptyset$

Now notice that $(t_i \cap L) \cap (t_j \cap L) \neq \emptyset$ for all pairs i, j. Indeed, for sake of contradiction, suppose there exist two edges t_i and t_j for $i < j$ such that either $t_i \cap L$ lies completely to the left of $t_j \cap L$, or vice versa. See Fig. 6.

Case 1. Suppose $t_i \cap L$ lies completely to the left of $t_j \cap L$. Then the vertical line through v_j intersects the right side of t_i below v_j. Therefore, the right side of $\text{conv}(b_i \cup v_j)$ intersects L more to the right than the right side of $t_i = \text{conv}(b_i \cup v_i)$ does. This contradicts the definition of t_i and t_j.

Case 2. Suppose $t_i \cap L$ lies completely to the right of $t_j \cap L$. Then there exists a base b_s that has a point p on L between $t_i \cap L$ and $t_j \cap L$. Base b_s must

1. lie below v_i and v_j,
2. cross b_i and b_j, and
3. contain point p.

However, this is impossible by the following argument. Let l be the vertical line through p. Clearly, l intersects b_i and b_j. Since b_s lies below v_i and v_j, b_s must intersect b_j to the left of l and intersect b_i to the right of l. Since b_s intersects b_j to the left of l, the slope of b_s must be greater than the slope of b_j. However, since the slope of b_i is less than the slope of b_j, this implies that b_s cannot intersect b_i to the right of l. Hence, we have a contradiction (Fig. 7).
Since $(t_i \cap L) \cap (t_j \cap L) \neq \emptyset$ for every $i, j, \in \{1, 2, \ldots, k\}$, by Helly's theorem [6], $t_1, \ldots, t_k$ has a nonempty intersection on L. □

Notice that no k points in $V(H)$ have $c_k n$ bases in common. Indeed, otherwise the vertical line through any of these k points would intersect all $c_k n$ bases, and by Lemma 2.3 there would be k pairwise crossing bases. By Observation 2.4, we would have k strongly crossing edges.

Now let $G = (A \cup B, E)$ be a bipartite graph where $A = V(H)$ and $B = V^2(H)$, such that $(v, xy) \in E(G)$ if $\text{conv}(x \cup y \cup v) \in E(H)$ with base xy. Since G does not contain $K_{k,c_k n}$ as a subgraph, we can use the following well-known result of Kővári et al. [10].

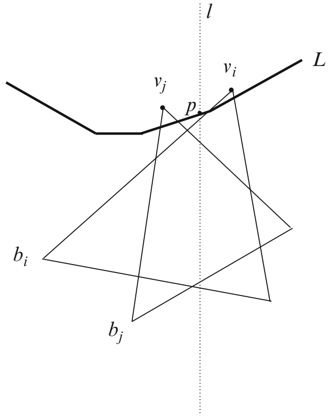


Fig. 7 Case 2

Theorem 5. *If $G = (A \cup B, E)$ is a bipartite graph with $|A| = n$ and $|B| = m$ containing no subgraph $K_{r,s}$ with the r vertices in A and the s vertices in B, then*

$$|E(G)| \leq (s-1)^{1/r} n m^{1-1/r} + (r-1)m.$$

By plugging the values $m = n^2, r = k, s = c_k n$ into Theorem 5, we obtain

$$\frac{|E(H)|}{2} \leq |E(G)| \leq O\left(n^{3-\frac{1}{k}}\right).$$

Hence,

$$|E(H)| \leq O\left(n^{3-\frac{1}{k}}\right).$$ □

2.1 Convex Geometric 3-Hypergraphs

In the case when the vertices are in convex position in the plane, extremal problems on geometric 3-hypergraphs become easier due to the linear ordering of its vertices. The proof of Observation 2.4 can be copied almost verbatim to conclude the following.

Observation 2.6. *Let $H = (V, E)$ be a geometric 3-hypergraph in the plane with vertices in convex position. Suppose H contains k edges of the form $t_i = \mathrm{conv}(x_i \cup y_i \cup z_i)$, such that the vertices $(x_1, \ldots, x_k, y_1, \ldots, y_k, z_1, \ldots, z_k)$ appear in clockwise order along the boundary of their convex hull. Then $t_1, \ldots, t_k$ are k strongly crossing edges.* □

Marcus and Klazar [9] extended the Marcus–Tardos theorem [12] by showing that the number of 1 entries in an r-dimensional $(0,1)$-matrix with side length n that avoids an r-dimensional permutation matrix is $O(n^{r-1})$. As pointed out by Marcus and Klazar, it is not difficult to modify their proof to obtain an $O(n^{r-1})$ bound on the number of edges in an ordered n-vertex r-uniform hypergraph that does not contain a fixed ordered matching. Hence, by Observation 2.6, we can conclude the following.

Theorem 7. *Let $H = (V, E)$ be a geometric 3-hypergraph in the plane with vertices in convex position. If H does not contain k strongly crossing edges, then $|E(H)| \leq c_k n^2$, where c_k is a constant that depends only on k.* □

3 Disjoint Edges in 3-Space

In this section, we will prove Theorem 5. Recall that two edges in a geometric graph are *parallel* if they are the opposite edges of a convex quadrilateral. Katchalski and Last [7] and Pinchasi [15] showed that all n-vertex geometric graphs with more than $2n - 2$ edges contain two parallel edges. By following Pinchasi's argument almost verbatim, one can prove the following.

Lemma 3.1. *Let G be a graph drawn on the unit sphere S with vertices represented as points such that no three lie on a great circle, and edges $uv \in E(G)$ are drawn as arcs along the great circle containing points u and v of length less than π (the shorter arc). We say that edges $e_1, e_2 \in E(G)$ are avoiding if the great circle supported by e_1 is disjoint to e_2, and the great circle supported by e_2 is disjoint from e_1. If $|E(G)| > 2n - 2$, then G contains two avoiding edges.* □

Proof of Theorem 1.5. Let $H = (V, E)$ be an n-vertex geometric 3-hypergraph in 3-space with no two disjoint edges. Fix a pair of vertices $u, v \in V(H)$, and just consider the edges $E_{uv} = \{t \in E(H) : u, v \text{ are vertices of } t\}$. We color $t \in E_{uv}$ *red* if all of the members of E_{uv} lie in one of the closed half-spaces generated by the plane supported by t. Notice that there are at most two red edges in E_{uv}. Repeat this procedure for each pair of vertices, which will leave us with at most n^2 red edges in the end. Color the remaining edges blue, and let $d_b(v)$ denote the number of blue edges incident to v. Then we have

$$\sum_{v \in V(H)} d_b(v) \geq 3E(H) - 3n^2.$$

Therefore, there exists a vertex v incident to at least $(3|E(H)| - 3n^2)/n$ blue edges. Now consider a small two-dimensional sphere S^2 centered at v. Then the intersection of S^2 and the blue edges incident to v forms a graph G with at most n vertices and at least $(3E(H) - 3n^2)/n$ edges.

If $(3|E(H)| - 3n^2)/n > 2n - 2$, then by Lemma 3.1 we know that G contains two avoiding edges xy and wz. Let h be the plane supported by the blue edge $\text{conv}(w \cup z \cup v) \in E(H)$. Then the blue edge $\text{conv}(x \cup y \cup v)$ must lie in one of the closed half-spaces generated by the plane h. Since $\text{conv}(w \cup z \cup v)$ is blue, there must be

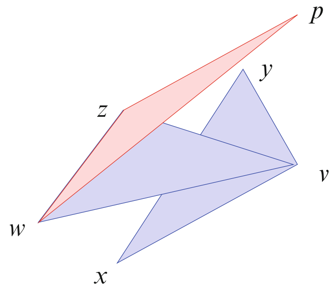


Fig. 8 Disjoint edges $\mathrm{conv}(w \cup z \cup p)$ and $\mathrm{conv}(x \cup y \cup v)$

a red edge $\mathrm{conv}(w \cup z \cup p)$ such that h separates it from $\mathrm{conv}(x \cup y \cup v)$. Hence, $\mathrm{conv}(x \cup y \cup v)$ and $\mathrm{conv}(w \cup z \cup p)$ are disjoint, and we have a contradiction. See Fig. 8. Therefore, $(3|E(H)| - 3n^2)/n \leq 2n - 2$, which implies $|E(H)| \leq O(n^2)$. □

4 Remarks

By applying the abstract crossing lemma (see [18]) to Theorem 2, every n-vertex geometric 3-hypergraph H in the plane has either $O(n^2)$ edges or $\Omega(|E(H)|^7/n^{12})$ triples that have a point in common. In the latter case, by the fractional Helly theorem [8], this implies one can always find a point inside at least $\Omega(|E(H)|^5/n^{12})$ edges of H. However, this is not as strong as the

$$\Omega\left(\frac{|E(H)|^3}{n^6 \log^2 n}\right)$$

bound obtained by Nivasch and Sharir [13].

Acknowledgements The author gratefully acknowledges the support from the Swiss National Science Foundation, Grant no. 200021-125287/1.

References

1. E. Ackerman, On the maximum number of edges in topological graphs with no four pairwise crossing edges, in *Proceedings of the Twenty-Second Annual Symposium on Computational Geometry*, Sedona, AZ, USA, 05–07 June 2006. SCG '06 (ACM, New York, 2006), pp. 259–263
2. J. Akiyama, N. Alon, Disjoint simplices and geometric hypergraphs, in *Proceedings of the Third international Conference on Combinatorial Mathematics*, ed. by G.S. Bloom, R.L. Graham, J. Malkevitch. New York City (New York Academy of Sciences, New York, 1989), pp. 1–3
3. N. Alon, I. Bárány, Z. Füredi, D.J. Kleitman, Point selections and weak ε-nets for convex hulls. Combin. Probab. Comput. **1**, 189–200 (1992)

4. K. Borsuk, Drei Sätze ber die n-dimensionale euklidische Sphäre. Fund. Math. **20**, 177–190 (1933)
5. T.K. Dey, J. Pach, Extremal problems for geometric hypergraphs, in *Discrete and Computational Geometry*, 19, 473-484 (1998)
6. E. Helly, Über Mengen konvexer Körper mit gemeinschaftlichen Punkten. Jber. Deutsch. Math. Vereinig. **32**, 175–176 (1923)
7. M. Katchalski, L. Last, On geometric graphs with no two edges in convex position. Discrete Comput. Geom. **19**(3), Special Issue, 399–404 (1998)
8. M. Katchalski, A. Liu, A problem of geometry in $\mathbb{R}^n$. Proc. Am. Math. Soc. **75**, 284–288 (1979)
9. M. Klazar, A. Marcus, Extensions of the linear bound in the Füredi–Hajnal conjecture. Adv. Appl. Math. **38**(2), 258–266 (2006)
10. T. Kővári, V. Sós, P. Turán, On a problem of K. Zarankiewicz. Coll. Math. **3**, 50–57 (1954)
11. J. Matoušek, *Lectures on Discrete Geometry* (Springer-Verlag, New York, 2002)
12. A. Marcus, G. Tardos, Excluded permutation matrices and the Stanley–Wilf conjecture. J. Combin. Theor. Ser. A **107**(1), 153–160 (2004)
13. G. Nivasch, M. Sharir, Note: Eppstein's bound on intersecting edges revisited. J. Comb. Theor. Ser. A **116**(2), 494–497 (2009)
14. J. Pach, J. Törőcsik, Some geometric applications of Dilworth's theorem, in *Proceedings of the Ninth Annual Symposium on Computational Geometry*, San Diego, CA, 18–21 May 1993. SCG '93 (ACM, New York, 1993), pp. 264–269
15. R. Pinchasi, Geometric graphs with no two parallel edges. Combinatorica **28**, 1, 127–130 (2008)
16. G. Tóth, Note on geometric graphs. J. Comb. Theor. Ser. A **89**(1), 126–132 (2000)
17. P. Valtr, Graph drawings with no k pairwise crossing edges, in *Graph Drawing* (Rome). Lecture Notes in Computer Science, vol. 1353, 1997, pp. 205–218
18. U. Wagner, k-Sets and k-Facets, discrete and computational geometry—20 years later, in *Contemporary Mathematics*, vol. 453, ed. by E. Goodman, J. Pach, R. Pollack (American Mathematical Society, New York, 2008)
19. R.T. Živaljević, S.T. Vrećica, The colored Tverberg's problem and complexes of injective functions. J. Comb. Theor. Ser. A **61**(2), 309–318 (1992)

Favorite Distances in High Dimensions

Konrad J. Swanepoel

Abstract Let S be a set of n points in $\mathbb{R}^d$. Assign to each $\boldsymbol{x} \in S$ an arbitrary distance $r(\boldsymbol{x}) > 0$. Let $e_r(\boldsymbol{x}, S)$ denote the number of points in S at distance $r(\boldsymbol{x})$ from $\boldsymbol{x}$. Avis, Erdős, and Pach (1988) introduced the extremal quantity $f_d(n) = \max \sum_{\boldsymbol{x} \in S} e_r(\boldsymbol{x}, S)$, where the maximum is taken over all n-point subsets S of $\mathbb{R}^d$ and all assignments $r\colon S \to (0, \infty)$ of distances.

We give a quick derivation of the asymptotics of the error term of $f_d(n)$ using only the analogous asymptotics of the maximum number of unit distance pairs in a set of n points:

$$f_d(n) = \left(1 - \frac{1}{\lfloor d/2 \rfloor}\right) n^2 + \begin{cases} \Theta(n) & \text{if } d \text{ is even,} \\ \Theta((n/d))^{4/3}) & \text{if } d \text{ is odd.} \end{cases}$$

The implied constants are absolute. This improves on previous results of Avis, Erdős, and Pach (1988) and Erdős and Pach (1990).

Then we prove a stability result for $d \geq 4$, asserting that if (S, r) with $|S| = n$ satisfies $e_r(S) = f_d(n) - o(n^2)$, then, up to $o(n)$ points, S is a Lenz construction with r constant. Finally, we use stability to show that for n sufficiently large (depending on d), the pairs (S, r) that attain $f_d(n)$ are up to scaling exactly the Lenz constructions that maximize the number of unit distance pairs with $r \equiv 1$, with some exceptions in dimension 4.

Analogous results hold for the furthest-neighbor digraph, where r is fixed to be $r(\boldsymbol{x}) = \max_{\boldsymbol{y} \in S} |\boldsymbol{xy}|$ for $\boldsymbol{x} \in S$.

K.J. Swanepoel (✉)
Department of Mathematics, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK
e-mail: k.swanepoel@lse.ac.uk

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_27,

1 Introduction

Denote the d-dimensional Euclidean space by $\mathbb{R}^d$, and the Euclidean distance between points $\boldsymbol{x}$ and $\boldsymbol{y}$ by $|\boldsymbol{xy}|$. Let S be a set of n points in $\mathbb{R}^d$. Let $r\colon S \to (0,\infty)$ be a choice of a positive number for each point in S. Define the *favorite distance digraph on S determined by r* to be the directed graph $\vec{G}_r(S) = (S, \vec{E}_r(S))$ on the set S, where

$$\vec{E}_r(S) := \{(\boldsymbol{x},\boldsymbol{y}) : \boldsymbol{x},\boldsymbol{y} \in S \text{ and } |\boldsymbol{xy}| = r(\boldsymbol{x})\}.$$

Write $e_r(S) := \left|\vec{E}_r(S)\right|$. Define

$$f_d(n) := \max\left\{e_r(S) : S \subset \mathbb{R}^d, |S| = n \text{ and } r\colon S \to (0,\infty)\right\}.$$

Define $D = D_S\colon S \to (0,\infty)$ by

$$D(\boldsymbol{x}) := \max\{|\boldsymbol{xs}| : \boldsymbol{s} \in S\}.$$

Then $\vec{G}_D(S)$ is called the *furthest-neighbor digraph* of S. Define

$$g_d(n) := \max\left\{e_D(S) : S \subset \mathbb{R}^d, |S| = n\right\}.$$

Clearly, $g_d(n) \leq f_d(n)$. In fact, $g_d(n) \sim f_d(n) \sim (1 - 1/\lfloor d/2 \rfloor)n^2$ for any fixed $d \geq 4$ as $n \to \infty$ [3, 11]. A set S of n points and a function $r\colon S \to (0,\infty)$ define an *extremal favorite distance digraph* if $e_r(S) = f_d(|S|)$. Likewise, S defines an *extremal furthest-neighbor digraph* if $e_D(S) = g_d(|S|)$.

1.1 Overview

In this chapter, we prove a structure theorem for extremal favorite distance digraphs and furthest-neighbor digraphs (Theorem B) for dimension $d \geq 4$. This structure theorem follows from a stability result describing the pairs (S, r) for which $e_r(S)$ is close to $f_d(n)$ (Theorem C). In Sect. 2, we start off with an easy derivation of the optimal asymptotics of the error term of $f_d(n)$ (Theorem A). This simple proof introduces the basic approach used in this chapter. Section 3 gives a description of the Lenz configurations and formally states Theorem B. Then we state Theorem C in Sect. 4. Section 5 contains the proof of Theorem C and Sect. 6 the proof of Theorem B.

Note that we only consider dimensions $d \geq 4$ in this chapter. For lower dimensions, we only make the following remarks. The current best estimates

$$\frac{n^2}{4} + \frac{5n}{2} - 6 \le f_3(n) \le \frac{n^2}{4} + \frac{5n}{2} + 6$$

for large n can be found in another paper [13]. Csizmadia [7] determined $g_3(n)$ exactly for large n. In dimension 2, a construction gives $f_2(n) = \Omega(n^{4/3})$ [5, p. 187], while the best-known upper bound $f_2(n) = O(n^{15/11+\varepsilon})$ is due to Aronov and Sharir [1]. Avis [2] and Edelsbrunner and Skiena [8] determined $g_2(n)$ exactly.

Throughout this chapter, $[k]$ denotes the set $\{1,2,\dots,k\}$, $\binom{S}{2}$ the set of unordered pairs of elements of S, K_p the complete graph on p vertices, and $K_p(t)$ the complete p-partite graph with t elements in each class.

2 Asymptotics

The problem of determining $f_d(n)$ and $g_d(n)$ was originally introduced by Avis, Erdős, and Pach [3]. They determined $f_d(n)$ asymptotically for even $d \ge 4$. Erdős and Pach [11] finished off the case of odd $d \ge 5$.

Theorem 1 (Avis–Erdős–Pach [3], Erdős–Pach [11]). *For any $d \ge 4$,*

$$f_d(n) = \left(1 - \frac{1}{\lfloor d/2 \rfloor} + o(1)\right) n^2.$$

We note that for even dimensions $d \ge 4$, the error term in [3] is $O(n^{2-\varepsilon})$ for some $\varepsilon > 0$ independent of d. The lower bound is obtained from the corresponding lower bound for the maximum number $u_d(n)$ of unit distance pairs in a set of n points in $\mathbb{R}^d$ (the *Lenz construction* [9]; see Sect. 3). For any set $S \subset \mathbb{R}^d$ of n points, let

$$u(S) := |\{\{\boldsymbol{x},\boldsymbol{y}\} \ : \ \boldsymbol{x},\boldsymbol{y} \in S \text{ and } |\boldsymbol{xy}| = 1\}|$$

and set

$$u_d(n) := \max\left\{u(S) \ : \ S \subset \mathbb{R}^d \text{ and } |S| = n\right\}.$$

Clearly, $f_d(n) \ge 2u_d(n)$. Similarly, $g_d(n) \ge 2M_d(n)$, where $M_d(n)$ is the maximum number of diameter pairs in a set of n points in $\mathbb{R}^d$, defined by setting

$$M(S) := |\{\{\boldsymbol{x},\boldsymbol{y}\} \ : \ \boldsymbol{x},\boldsymbol{y} \in S \text{ and } |\boldsymbol{xy}| = \operatorname{diam}(S)\}|$$

and

$$M_d(n) := \max\left\{M(S) \ : \ S \subset \mathbb{R}^d, |S| = n\right\}.$$

We show that the determination of $f_d(n)$ and $g_d(n)$ in effect reduces to the unit distance problem when $d \ge 4$. A first indication of this is a simple derivation of an asymptotic upper bound for $f_d(n)$ (Theorem A below) using only the analogous upper bounds for $u_d(n)$ stated in the following theorem.

Theorem 2 (Erdős [10], Erdős–Pach [11]). *There exist constants $c_1, c_2 > 0$ such that for each $d \geq 4$ and all $n \in \mathbb{N}$,*

$$u_d(n) \leq \frac{1}{2}\left(1 - \frac{1}{\lfloor d/2 \rfloor}\right) n^2 + \begin{cases} c_1 n & \text{if } d \text{ is even,} \\ c_2 \left(\frac{n}{d}\right)^{4/3} & \text{if } d \text{ is odd.} \end{cases}$$

The above bounds are tight up to the values of c_1 and c_2 [11]. In fact, Erdős proved that for even $d \geq 4$ and sufficiently large n,

$$\frac{1}{2}\left(1 - \frac{2}{d}\right) n^2 + n - \frac{d}{2} \leq u_d(n) \leq \frac{1}{2}\left(1 - \frac{2}{d}\right) n^2 + n.$$

However, in the proof of the next theorem, we need a bound that holds for all $n \in \mathbb{N}$. Since $f_d(n) \geq 2u_d(n)$, the bounds in the next theorem are also tight up to the values of the constants.

Theorem A. *With the same constants $c_1, c_2 > 0$ as in Theorem 2, for each $d \geq 4$ and all $n \in \mathbb{N}$,*

$$f_d(n) \leq \left(1 - \frac{1}{\lfloor d/2 \rfloor}\right) n^2 + \begin{cases} 2c_1 n & \text{if } d \text{ is even,} \\ 2c_2 \left(\frac{n}{d}\right)^{4/3} & \text{if } d \text{ is odd.} \end{cases}$$

Proof. Let $S \subset \mathbb{R}^d$ be an arbitrary set of n points and $r\colon S \to (0,\infty)$ any function that assigns a positive real number to each point in S. We next introduce notation and terminology that will also be used in later proofs. We first decompose $\vec{G}_r(S)$ into two ordinary graphs. Let $G_r^1(S) = (S, E_r^1)$ be the graph of *single edges*, where

$$E_r^1 := \Big\{ \{\boldsymbol{x}, \boldsymbol{y}\} \;:\; (\boldsymbol{x}, \boldsymbol{y}) \in \vec{E}_r(S), (\boldsymbol{y}, \boldsymbol{x}) \notin \vec{E}_r(S) \Big\}.$$

Let $G_r^2(S) = (S, E_r^2)$ be the graph of *double edges*, where

$$E_r^2 := \Big\{ \{\boldsymbol{x}, \boldsymbol{y}\} \;:\; (\boldsymbol{x}, \boldsymbol{y}), (\boldsymbol{y}, \boldsymbol{x}) \in \vec{E}_r(S) \Big\}.$$

Write the connected components of $G_r^2(S)$ as $G_r^2[S_i]$ for $i \in [k]$, where $\{S_1, \dots, S_k\}$ partitions S. For subsets $A, B \subseteq S$, let

$$\vec{E}_r(A, B) := \Big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \vec{E}_r(S) \;:\; \boldsymbol{x} \in A, \boldsymbol{y} \in B \Big\}$$

and $e_r(A, B) := \Big|\vec{E}_r(A, B)\Big|$. Write $n_i := |S_i|$ for each $i \in [k]$.

The proof is based on the following two simple facts.

1. Each $G_r^2[S_i]$ is a scaling of a unit distance graph. Therefore, $e_r(S_i) \leq u_d(n_i)$.
2. There can only be single edges between different S_i. Consequently,

$$e_r(S_i,S_j)+e_r(S_j,S_i) \leq n_i n_j \quad \text{for any distinct } i,j.$$

The proof is finished by a calculation. Note that

$$\begin{aligned} e_r(S) &= \sum_{i=1}^{k} e_r(S_i) + \sum_{\{i,j\}\in\binom{[k]}{2}} (e_r(S_i,S_j)+e_r(S_j,S_i)) \\ &\leq \sum_{i=1}^{k} 2u_d(n_i) + \sum_{\{i,j\}\in\binom{[k]}{2}} n_i n_j. \end{aligned}$$

Now fix S and r so that $f_d(n) = e_r(S)$, and apply Theorem 2 to obtain for odd dimensions $d \geq 5$ that

$$\begin{aligned} f_d(n) &\leq \sum_{i=1}^{k}\left(2u_d(n_i) - \frac{1}{2}n_i^2\right) + \sum_{i=1}^{k}\frac{1}{2}n_i^2 + \sum_{\{i,j\}\in\binom{[k]}{2}} n_i n_j \\ &= \sum_{i=1}^{k}\left(2u_d(n_i) - \frac{1}{2}n_i^2\right) + \frac{1}{2}n^2 \\ &\leq \sum_{i=1}^{k}\left(\left(1-\frac{1}{\lfloor d/2\rfloor}\right)n_i^2 + 2c_2\left(\frac{n_i}{d}\right)^{4/3} - \frac{1}{2}n_i^2\right) + \frac{1}{2}n^2 \\ &= \sum_{i=1}^{k}\left(\left(\frac{1}{2}-\frac{1}{\lfloor d/2\rfloor}\right)n_i^2 + 2c_2\left(\frac{n_i}{d}\right)^{4/3}\right) + \frac{1}{2}n^2 \\ &\leq \left(\frac{1}{2}-\frac{1}{\lfloor d/2\rfloor}\right)\left(\sum_{i=1}^{k} n_i\right)^2 + 2c_2\left(\frac{\sum_{i=1}^{k} n_i}{d}\right)^{4/3} + \frac{1}{2}n^2 \\ &= \left(1-\frac{1}{\lfloor d/2\rfloor}\right)n^2 + 2c_2\left(\frac{n}{d}\right)^{4/3}, \end{aligned}$$

where the last inequality follows from the inequality

$$\sum_{i=1}^{k} n_i^{\alpha} \leq \left(\sum_{i=1}^{k} n_i\right)^{\alpha} \qquad \text{for all } n_i \geq 0 \text{ and } \alpha \geq 1, \tag{1}$$

which is easily seen to be true (for example, from Minkowski's inequality). The calculation for even values of $d \geq 4$ is similar. □

3 Extremal Configurations

By a *Lenz configuration for distance* $\lambda > 0$, we mean a finite set of the following type [4, 9].

If $d \geq 4$ is even, let $p = d/2$ and consider any orthogonal decomposition $\mathbb{R}^d = V_1 \oplus \cdots \oplus V_p$ with all V_i two-dimensional. In each V_i, let C_i be the circle with center at the origin $\boldsymbol{o}$ and radius r_i, such that $r_i^2 + r_j^2 = \lambda^2$ for all distinct i and j. When $d \geq 6$, this implies that each $r_i = \lambda/\sqrt{2}$. We call the p circles $(C_1, \dots, C_p)$ an *even-dimensional Lenz system*. Define an *even-dimensional Lenz configuration for the distance* λ to be any finite subset S of some translate $\boldsymbol{v} + \bigcup_{i=1}^{p} C_i$ of the circles. The *partition associated with the Lenz configuration* S is the partition induced by the circles, i.e., the p subsets $S_1, \dots, S_p$, where $S_i = S \cap (\boldsymbol{v} + C_i)$.

If $d \geq 5$ is odd, let $p = \lfloor d/2 \rfloor$, and consider any orthogonal decomposition $\mathbb{R}^d = V_1 \oplus \cdots \oplus V_p$ with V_1 three-dimensional and all other V_i $(i = 2, \dots, p)$ two-dimensional. Let Σ_1 be the 2-sphere in V_1 with center $\boldsymbol{o}$ and radius r_1, and for each $i = 2, \dots, p$, let C_i be the circle with center $\boldsymbol{o}$ and radius r_i, such that $r_i^2 + r_j^2 = \lambda^2$ for all distinct i, j. When $d \geq 7$, necessarily each $r_i = \lambda/\sqrt{2}$. We call the 2-sphere and $p-1$ circles $(\Sigma_1, C_2, \dots, C_p)$ an *odd-dimensional Lenz system*. We define an *odd-dimensional Lenz configuration for the distance* λ to be any finite subset of some translate $\boldsymbol{v} + \left(\Sigma_1 \cup \bigcup_{i=2}^{p} C_i\right)$ of the 2-sphere and circles. The *partition associated with the Lenz configuration* S is the partition induced by the 2-sphere and circles, i.e., the p subsets $S_1, \dots, S_p$, where $S_1 = S \cap \Sigma_1$ and for $i \geq 2$, $S_i = S \cap (\boldsymbol{v} + C_i)$. The following theorem states that the extremal sets for unit distances and for diameters are Lenz configurations, at least for a sufficiently large number of points.

Theorem 3 ([4, 12]). *For any $d \geq 4$, there exists $n_0 \in \mathbb{N}$ such that any set S for which $|S| = n \geq n_0$ and such that $u(S) = u_d(n)$ is a Lenz configuration for the distance* 1.

For any $d \geq 4$, there exists $n_0 \in \mathbb{N}$ such that any set S for which $|S| = n \geq n_0$ and such that $M(S) = M_d(n)$ is a Lenz configuration for the distance $\operatorname{diam}(S)$.

As a corollary of the main result of this chapter (Theorem C, in Sect. 4) we show that when $d \geq 4$, the extremal favorite distance digraphs (furthest-neighbor digraphs) are exactly the same as the sets for which $u_d(n)$ [$M_d(n)$ respectively] is maximized, for all sufficiently large n, depending on d, except when $d = 4$, where there is an exceptional construction for all sufficiently large $n \equiv 1 \pmod 8$.

Theorem B. *For any $d \geq 4$, there exists $n_0 \in \mathbb{N}$ such that the following holds.*

1. *Let $S \subset \mathbb{R}^d$ and a function $r\colon S \to (0, \infty)$ be given for which $|S| = n \geq n_0$ and $e_r(S) = f_d(n)$. Then $r \equiv c$ for some $c > 0$ and S is a Lenz configuration for the distance c, except when $d = 4$ and $n - 1$ is divisible by 8, where the following situation is also possible: For some $a \in S$ and $c > 0$, $S \setminus \{a\}$ is a Lenz configuration for the distance c on two circles C_1 and C_2 of equal radius $c/\sqrt{2}$, a is the common center of the two circles, $C_i \cap S$ consists of the vertices of $(n-1)/8$ squares inscribed in C_i $(i = 1, 2)$, and $r|_{S \setminus \{a\}} \equiv c$, $r(a) = c/\sqrt{2}$.*

2. *Let $S \subset \mathbb{R}^d$ be given for which $|S| = n \geq n_0$ and $e_D(S) = g_d(n)$. Then $r \equiv \mathrm{diam}(S)$ and S is a Lenz configuration for the distance* $\mathrm{diam}(S)$.

In particular, $f_d(n) = 2u_d(n)$ and $g_d(n) = 2M_d(n)$ for all $d \geq 4$ and $n \geq n_0(d)$.

Note that the exact values of $u_d(n)$ for even $d \geq 4$ and of $M_d(n)$ for all $d \geq 4$ are known, at least for sufficiently large n [4, 12]; see Lemmas 6.5 and 6.6 for some of these values.

4 Stability

The following theorem states that if the number of unit distance pairs of points from $S \subset \mathbb{R}^d$, where $n := |S|$ is sufficiently large, is within $o(n^2)$ within $o(n^2)$ of the maximum $u_d(n)$, then S is a Lenz configuration up to $o(n)$ points.

Theorem 4 ([12]). *For any $d \geq 4$ and $\varepsilon > 0$, there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for any set S with $|S| = n \geq n_0$ that satisfies*

$$u(S) > \frac{1}{2}\left(1 - \frac{1}{p} - \delta\right)n^2 \quad (\textit{where } p = \lfloor d/2 \rfloor),$$

there exists a subset $T \subseteq S$ such that $|T| < \varepsilon n$ and $S \setminus T$ is a Lenz configuration. Furthermore, the partition $S_1, \ldots, S_p$ of $S \setminus T$ associated with the Lenz configuration satisfies

$$\frac{n}{p} - \varepsilon n < |S_i| < \frac{n}{p} + \varepsilon n \quad \textit{for all } i \in [p].$$

The next theorem is an analogue of the above theorem for favorite distance digraphs.

Theorem C. *For any $d \geq 4$ and any $\varepsilon > 0$, there exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that for any $S \subset \mathbb{R}^d$ with $|S| = n \geq n_0$ and any $r\colon S \to (0, \infty)$ that satisfy*

$$e_r(S) > \left(1 - \frac{1}{p} - \delta\right)n^2 \quad (\textit{where } p = \lfloor d/2 \rfloor),$$

there exist $T \subseteq S$ and $c > 0$ such that $|T| < \varepsilon n$, $S \setminus T$ is a Lenz configuration with distance c, and $r|_{S \setminus T} \equiv c$. Furthermore, the partition $S_1, \ldots, S_p$ of $S \setminus T$ associated with the Lenz configuration satisfies

$$\frac{n}{p} - \varepsilon n < |S_i| < \frac{n}{p} + \varepsilon n \quad \textit{for all } i \in [p].$$

By applying Theorem 4, the above theorem is relatively easy to prove for $d \geq 6$ but, surprisingly, takes some work in the cases $d \in \{4, 5\}$. This is not so much because the Lenz construction is slightly more complicated in dimensions 4 and 5, but rather due to certain complications in the extremal theory of digraphs not shared by the extremal theory of ordinary graphs [6].

5 Proof of Theorem C

Let $d \geq 4$ and $\varepsilon > 0$ be given. Without loss of generality, $\varepsilon < \frac{1}{20}$. Let $p = \lfloor d/2 \rfloor$. We take $\delta > 0$ to be sufficiently small depending only on ε and d. In particular, we need

- $\delta < \varepsilon^2/144$ and,
- after an application of stability for unit distances (Theorem 4), we may also assume that δ has been chosen so that for some $N \in \mathbb{N}$ (which we now fix), for any $S \subset \mathbb{R}^d$ with $|S| = n \geq N$, if $u(S) > \frac{1}{2}\left(1 - \frac{1}{p} - 32\delta\right)n^2$, then for some $T \subseteq S$ of size $|T| < \varepsilon n/3$, $S \setminus T$ is a Lenz configuration such that the number of elements in each part of the associated partition is in the interval $((1/p - \varepsilon/3)n, (1/p + \varepsilon/3)n)$.

We also take $n_0 \in \mathbb{N}$ sufficiently large depending only on ε, d, and δ, as follows. We need

- $n_0 > 9/\delta$ and $n_0 \geq 4N$,
- n_0 to be sufficiently large such that for all $n \geq n_0$, $f_3(n-2) + 4n < \left(\frac{1}{2} - \delta\right)n^2$ (by Avis–Erdős–Pach [3] $f_3(n) = \frac{n^2}{4} + O(n^{2-c})$; in [13], we show $f_3(n) \leq \frac{n^2}{4} + \frac{5n}{2} + 6$ for sufficiently large n),
- $n_0 > (2c_2/\delta)^{3/2} d^{-2}$, where c_2 is the constant from Theorems 2 and A,
- n_0 to be sufficiently large such that for all $n \geq \varepsilon n_0/4$, $f_5(n) < \left(\frac{1}{2} + \delta\right)n^2$ (Theorem 1), and
- n_0 to be sufficiently large such that the Erdős–Stone theorem guarantees that any graph on $n \geq n_0$ vertices and at least $\left(\frac{1}{3} + \delta\right)n^2$ edges contains a $K_4(p_0)$, where p_0 is a constant such that no orientation of $K_4(p_0)$ can be a subgraph of a favorite distance digraph in $\mathbb{R}^5$ (Lemmas 4 and 5 in [3]).

Choose $S \subset \mathbb{R}^d$ with $|S| = n \geq n_0$ and $r\colon S \to (0,\infty)$ such that $e_r(S) > \left(1 - \frac{1}{p} - \delta\right)n^2$. We continue with the notation established in the proof of Theorem A. Thus, let $G_r^1(S)$ be the graph of single edges and $G_r^2(S)$ the graph of double edges of $\vec{G}_r(S)$ with connected components $G_r^2[S_i]$ ($i \in [k]$). As before, $n_i = |S_i|$. Also, write

$$d_{i,j} := \frac{e_r(S_i, S_j)}{n_i n_j}$$

for distinct $i, j \in [k]$ and $\alpha_i := n_i/n$. Similar to the calculation in the proof of Theorem A,

$$
\begin{aligned}
e_r(S) &= \sum_{i=1}^{k} e_r(S_i) + \sum_{\{i,j\}\in\binom{[k]}{2}} (e_r(S_i,S_j)+e_r(S_j,S_i)) \\
&\le \sum_{i=1}^{k} 2u_d(n_i) + \sum_{\{i,j\}\in\binom{[k]}{2}} (d_{i,j}+d_{j,i})\alpha_i\alpha_j n^2 \\
&\le \sum_{i=1}^{k} \left(\left(1-\frac{1}{p}\right)(\alpha_i n)^2 + 2c_2\left(\frac{\alpha_i n}{d}\right)^{4/3}\right) + \sum_{\{i,j\}\in\binom{[k]}{2}} (d_{i,j}+d_{j,i})\alpha_i\alpha_j n^2.
\end{aligned}
$$

It is given that $e_r(S) > \left(1-\frac{1}{p}-\delta\right)n^2$. Therefore,

$$
\begin{aligned}
1-\frac{1}{p}-\delta &< \sum_{i=1}^{k}\left(1-\frac{1}{p}\right)\alpha_i^2 + 2c_2 d^{-4/3}n^{-2/3}\sum_{i=1}^{k}\alpha_i^{4/3} + \sum_{\{i,j\}\in\binom{[k]}{2}} (d_{i,j}+d_{j,i})\alpha_i\alpha_j \\
&\le \left(1-\frac{1}{p}\right)\sum_{i=1}^{k}\alpha_i^2 + \delta\left(\sum_{i=1}^{k}\alpha_i\right)^{4/3} + \sum_{\{i,j\}\in\binom{[k]}{2}} (d_{i,j}+d_{j,i})\alpha_i\alpha_j \\
&\qquad \text{(since } n \text{ is sufficiently large and using (1))} \\
&= 1-\frac{1}{p}+\delta - \sum_{\{i,j\}\in\binom{[k]}{2}} \left(2\left(1-\frac{1}{p}\right)-d_{i,j}-d_{j,i}\right)\alpha_i\alpha_j. \qquad (2)
\end{aligned}
$$

Therefore,

$$
\sum_{\{i,j\}\in\binom{[k]}{2}} \left(2\left(1-\frac{1}{p}\right)-d_{i,j}-d_{j,i}\right)\alpha_i\alpha_j < 2\delta. \qquad (3)
$$

As noted in the proof of Theorem A, there are no double edges between S_i and S_j when $i \ne j$. Consequently, $d_{i,j}+d_{j,i} \le 1$, and therefore,

$$
\sum_{\{i,j\}\in\binom{[k]}{2}} \left(1-\frac{2}{p}\right)\alpha_i\alpha_j < 2\delta.
$$

Assume for the moment that $d \ge 6$. Then $p \ge 3$, and hence $\sum_{\{i,j\}} \alpha_i\alpha_j < 6\delta$. Substituting back into (2), we obtain

$$
1-\frac{1}{p}-\delta < \left(1-\frac{1}{p}\right)\sum_{i=1}^{k}\alpha_i^2 + \delta + 6\delta,
$$

which gives

$$
\sum_{i=1}^{k}\alpha_i^2 > \frac{1-\frac{1}{p}-8\delta}{1-\frac{1}{p}} \ge 1-12\delta.
$$

Since $\sum_{i=1}^{k}\alpha_i = 1$, it follows that $\alpha_i > 1-12\delta$ for some $i \in [k]$. Without loss of generality, $\alpha_1 > 1-12\delta$. Calculating again,

$$\begin{aligned}\left(1-\frac{1}{p}-\delta\right)n^2 < e_r(S) &= e_r(S_1)+\sum_{i=2}^{k} e_r(S_i)+\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j n^2\\ &< e_r(S_1)+\sum_{i=2}^{n}(\alpha_i n)^2+6\delta n^2\\ &\le e_r(S_1)+\left(\sum_{i=2}^{n}\alpha_i\right)^2 n^2+6\delta n^2 \quad \text{(by (1))}\\ &< e_r(S_1)+(12\delta)^2 n^2+6\delta n^2, \end{aligned} \tag{4}$$

and assuming after scaling that $r|_{S_1} \equiv 1$, we obtain

$$2u(S) \ge e_r(S_1) > \left(1-\frac{1}{p}-7\delta-(12\delta)^2\right)n^2 > \left(1-\frac{1}{p}-32\delta\right)n^2.$$

By the choice of δ and n_0, the proof is concluded by an application of Theorem 4. This establishes the theorem for all dimensions $d \ge 6$.

The remaining cases are $d=4$ and $d=5$. The four-dimensional case of the theorem is implied by the five-dimensional case. In fact, the theorem for $d=5$ implies that when $S \subset \mathbb{R}^5$ is contained in an affine hyperplane H, then for some $T \subseteq S$ with $|T| < \varepsilon|S|$, $S\setminus T$ is the intersection of a five-dimensional Lenz configuration with H. Such an intersection is clearly either a four-dimensional Lenz configuration or becomes three-dimensional after removing at most two points. In the latter case,

$$e_r(S) \le f_3(n-2)+2(n-2)+2(n-1) < \left(\frac{1}{2}-\delta\right)n^2$$

by the choice of n_0. Thus, the former case necessarily occurs.

For the remainder of the proof, assume that $d=5$. Then $p=2$ and (3) can now be written as

$$\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j < \sum_{\{i,j\}\in\binom{[k]}{2}}(d_{i,j}+d_{j,i})\alpha_i\alpha_j+2\delta.$$

Thus, the graph of single edges $G_r^1(S)$ is almost the complete k-partite graph with classes $S_1,\dots,S_k$. We next apply the Erdős–Stone theorem to show that one of the S_i is large in the sense that $|S_i| = \Omega(n)$. [We have no control over k yet, and have to eliminate the possibility that k is large with each S_i small, which would imply that $\vec{G}_r(S)$ is close to a tournament—a case which would be difficult to handle geometrically.] Since n_0 and p_0 were chosen so that $G_r^1(S)$ does not contain a copy of $K_4(p_0)$, the Erdős–Stone theorem gives for sufficiently large n that

$$\left(\frac{1}{3}+\delta\right)n^2 > \left|E(G_r^1)\right| = \sum_{\{i,j\}\in\binom{[k]}{2}} (d_{i,j}+d_{j,i})\alpha_i\alpha_j n^2.$$

Therefore, $\sum_{\{i,j\}} \alpha_i\alpha_j < \frac{1}{3}+3\delta$, and

$$\sum_{i=1}^{k}\alpha_i^2 = \left(\sum_{i=1}^{k}\alpha_i\right)^2 - 2\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j > 1-2\left(\frac{1}{3}+3\delta\right) = \frac{1}{3}-6\delta.$$

It follows that for some $i\in[k]$, $\alpha_i > \frac{1}{3}-6\delta$. Without loss of generality, $\alpha_1 > \frac{1}{3}-6\delta > \frac{1}{4}$. Thus, $|S_1| > n/4$. This enables us to show next that S_1 is almost a Lenz configuration. Suppose to the contrary that $e_r(S_1) \le \left(\frac{1}{2}-32\delta\right)(\alpha_1 n)^2$. Starting off as in (4), an application of Theorem 2 now gives the following:

$$\begin{aligned}\left(\frac{1}{2}-\delta\right)n^2 &< e_r(S) = e_r(S_1)+\sum_{i=2}^{k}e_r(S_i)+\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j n^2\\ &< \left(\frac{1}{2}-32\delta\right)(\alpha_1 n)^2+\sum_{i=2}^{k}\left(\frac{1}{2}(\alpha_i n)^2+2c_2\left(\frac{\alpha_i n}{5}\right)^{4/3}\right)\\ &\quad+\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j n^2.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{1}{2}-\delta &< \frac{1}{2}\sum_{i=1}^{k}\alpha_i^2-32\delta\alpha_1^2+\sum_{i=2}^{k}\frac{2c_2}{5^{4/3}n^{2/3}}\alpha_i^{4/3}+\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j\\ &< \frac{1}{2}\sum_{i=1}^{k}\alpha_i^2-32\delta\alpha_1^2+\delta\sum_{i=2}^{k}\alpha_i^{4/3}+\sum_{\{i,j\}\in\binom{[k]}{2}}\alpha_i\alpha_j \quad \text{for } n \text{ sufficiently large}\\ &= \frac{1}{2}-32\delta\alpha_1^2+\delta\sum_{i=2}^{k}\alpha_i^{4/3} < \frac{1}{2}-32\delta\alpha_1^2+\delta\left(\sum_{i=2}^{k}\alpha_i\right)^{4/3} \quad \text{by (1)}\\ &< \frac{1}{2}-32\delta\alpha_1^2+\delta < \frac{1}{2}-\delta\end{aligned}$$

since $\alpha_1 > 1/4$. This contradiction gives (after scaling so that $r|_{S_1}\equiv 1$) that

$$2u(S_1) = e_r(S_1) > \left(\frac{1}{2}-32\delta\right)(\alpha_1 n)^2.$$

Since $|S_1| = \alpha_1 n > n_0/4 \geq N$ and by the choice of δ and n_0, Theorem 4 gives a $T \subset S_1$ with $|T| < \varepsilon |S_1|/3 \leq \varepsilon n/3$ such that $S_1 \setminus T$ is a Lenz configuration for the distance 1. Thus, we may write $\mathbb{R}^5 = V_1 \oplus V_2$ with $\dim V_1 = 3$ and $\dim V_2 = 2$ such that $S_1 \setminus T \subset \Sigma_1 \cup C_2$, where Σ_1 is a 2-sphere in V_1 with center $\boldsymbol{o}$ and radius r_1, and C_2 is a circle in V_2 with center $\boldsymbol{o}$ and radius r_2, where $r_1^2 + r_2^2 = 1$. After possibly replacing T by a subset, we may assume without loss of generality that $T \cap (\Sigma_1 \cup C_2) = \varnothing$. Also, then $|S_1 \cap \Sigma_1|, |S_1 \cap C_2| < (\frac{1}{2} + \frac{2\varepsilon}{3})|S_1|$. Since $S_1 \setminus T$ is a Lenz configuration and $|T| < \varepsilon n/3$, the proof would be finished if we can show that $|S \setminus S_1| < 2\varepsilon n/3$.

To this end, we will partition $S \setminus S_1$ into two parts $X \cup Y$, and estimate $e_r(S)$ from above by breaking it up as follows:

$$\begin{aligned} e_r(S) = e_r(S_1) + e_r(X) + e_r(S_1, X) + e_r(X, S_1) \\ + e_r(Y) + e_r(S_1 \cup X, Y) + e_r(Y, S_1 \cup X). \end{aligned} \tag{5}$$

Write $n_1 := |S_1|$. To define Y, we introduce the following notion. A circle C on Σ_1 is said to be *rich* if $|S_1 \cap C| \geq n_1/8$. If there are at least 5 rich circles on Σ_1, inclusion–exclusion gives (since two circles intersect in at most two points) that

$$\left(\frac{1}{2} + \frac{2\varepsilon}{3}\right) n_1 > |S_1 \cap \Sigma_1| \geq 5\frac{n_1}{8} - 2\binom{5}{2},$$

which leads to a contradiction for sufficiently large $n_1 > n/4$.

Therefore, there are at most 4 rich circles on Σ_1. Let Y be the set of all points in $(S \setminus S_1) \cap V_1$ that are equidistant to some rich circle. Let $X := S \setminus (S_1 \cup Y)$. Write $x := |X|$ and $y := |Y|$. Note that Y can be covered by 4 lines, since the points in the three-dimensional V_1 that are equidistant to some circle all lie on a line. Since any point is equidistant to at most 2 points on a line, we have $e_r(\boldsymbol{x}, Y) \leq 8$ for all $\boldsymbol{x} \in S$; hence,

$$\begin{aligned} e_r(Y) + e_r(S_1 \cup X, Y) + e_r(Y, S_1 \cup X) &\leq 8y + 8(n_1 + x) + y(n_1 + x) \\ &= 8n + yn_1 + yx. \end{aligned} \tag{6}$$

To bound $e_r(S_1, X) + e_r(X, S_1)$ from above, we estimate $e_r(S_1, \boldsymbol{x}) + e_r(\boldsymbol{x}, S_1)$ for $\boldsymbol{x} \in X$. If $r(\boldsymbol{x}) = 1$, then $e_r(S_1, \boldsymbol{x}) + e_r(\boldsymbol{x}, S_1) = 0$ since S_1 is the vertex set of a connected component of the graph $G_r^2(S)$ of double edges and $\boldsymbol{x} \notin S_1$. Thus, we may assume without loss of generality that $r(\boldsymbol{x}) \neq 1$.

If $e_r(S_1 \cap C_2, \boldsymbol{x}) + e_r(\boldsymbol{x}, S_1 \cap C_2) \leq 4$, then

$$\begin{aligned} &e_r(S_1, \boldsymbol{x}) + e_r(\boldsymbol{x}, S_1) \\ &\quad = e_r(S_1 \cap C_2, \boldsymbol{x}) + e_r(S_1 \setminus C_2, \boldsymbol{x}) + e_r(\boldsymbol{x}, S_1 \cap C_2) + e_r(\boldsymbol{x}, S_1 \setminus C_2) \\ &\quad \leq 4 + e_r(S_1 \setminus C_2, \boldsymbol{x}) + e_r(\boldsymbol{x}, S_1 \setminus C_2) \end{aligned}$$

$$\begin{aligned}
&\leq 4+|S_1\setminus C_2| = 4+|S_1\cap\Sigma_2|+|T|\\
&< 4+\left(\frac{1}{2}+\frac{2\varepsilon}{3}\right)n_1+\frac{\varepsilon}{3}n_1 = 4+\left(\frac{1}{2}+\varepsilon\right)n_1\\
&< \left(\frac{3}{4}+\varepsilon\right)n_1 \quad \text{for } n_1>n/4 \text{ sufficiently large.}
\end{aligned}$$

Otherwise, $e_r(S_1\cap C_2,\boldsymbol{x})+e_r(\boldsymbol{x},S_1\cap C_2)\geq 5$, and then either $e_r(S_1\cap C_2,\boldsymbol{x})\geq 3$ or $e_r(\boldsymbol{x},S_1\cap C_2)\geq 3$. In both cases, $\boldsymbol{x}$ is equidistant to C_2, which implies that $\boldsymbol{x}\in V_1$. Also,

$$e_r(S_1\cap C_2,\boldsymbol{x})+e_r(\boldsymbol{x},S_1\cap C_2) = |S_1\cap C_2| < \left(\frac{1}{2}+\frac{2\varepsilon}{3}\right)n_1.$$

If $\boldsymbol{x}\neq\boldsymbol{o}$, then the points on Σ_1 at distance 1 to $\boldsymbol{x}$ lie on a circle. Since $\boldsymbol{x}\notin Y$, this circle is not rich, giving $e_r(S_1\cap\Sigma_1,\boldsymbol{x})<n_1/8$. Similarly, $e_r(\boldsymbol{x},S_1\cap\Sigma_1)<n_1/8$. Putting everything together, we obtain

$$\begin{aligned}
&e_r(S_1,\boldsymbol{x})+e_r(\boldsymbol{x},S_1)\\
&\quad= e_r(S_1\cap C_2,\boldsymbol{x})+e_r(\boldsymbol{x},S_1\cap C_2)+e_r(S_1\cap\Sigma_1,\boldsymbol{x})+e_r(\boldsymbol{x},S_1\cap\Sigma_1)\\
&\qquad\quad +e_r(T,\boldsymbol{x})+e_r(\boldsymbol{x},T)\\
&\quad< \left(\frac{1}{2}+\frac{2\varepsilon}{3}\right)n_1+\frac{n_1}{8}+\frac{n_1}{8}+\frac{\varepsilon}{3}n_1\\
&\quad= \left(\frac{3}{4}+\varepsilon\right)n_1.
\end{aligned}$$

We have shown that for all $\boldsymbol{x}\in X\setminus\{\boldsymbol{o}\}$,

$$e_r(S_1,\boldsymbol{x})+e_r(\boldsymbol{x},S_1) < \left(\frac{3}{4}+\varepsilon\right)n_1.$$

Also, if $\boldsymbol{o}\in X$, then $e_r(S_1,\boldsymbol{o})+e_r(\boldsymbol{o},S_1)\leq|S_1|=n_1$. Sum over all $\boldsymbol{x}\in X$ to obtain

$$e_r(S_1,X)+e_r(X,S_1) < \left(\frac{3}{4}+\varepsilon\right)n_1x+\frac{1}{4}n_1. \tag{7}$$

Since $n_1>n/4$ is sufficiently large,

$$e_r(S_1) < \left(\frac{1}{2}+\delta\right)n_1^2. \tag{8}$$

Recall that we want to show that $|S\setminus S_1|=|X\cup Y|<2\varepsilon n/3$. Suppose that $x\geq\varepsilon n/4$. Since n is sufficiently large,

$$e_r(X) < \left(\frac{1}{2}+\delta\right)x^2. \tag{9}$$

Substitute (6)–(9) into (5) to obtain

$$\begin{aligned}\left(\frac{1}{2}-\delta\right)n^2 < e_r(S) &< \left(\frac{1}{2}+\delta\right)n_1^2 + \left(\frac{1}{2}+\delta\right)x^2 + \left(\frac{3}{4}+\varepsilon\right)n_1x + \frac{1}{4}n_1 \\ &\quad + 8n + yn_1 + yx \\ &= \left(\frac{1}{2}+\delta\right)(n_1+x+y)^2 - \left(\frac{1}{2}+\delta\right)y^2 - 2\delta yn_1 - 2\delta yx \\ &\quad - \left(\frac{1}{4}+2\delta-\varepsilon\right)n_1x + 8n + \frac{1}{4}n_1 \\ &< \left(\frac{1}{2}+\delta\right)n^2 - \left(\frac{1}{2}+\delta\right)y^2 - \left(\frac{1}{4}-\frac{1}{20}\right)\frac{n}{4}\cdot\frac{\varepsilon n}{4} + 9n,\end{aligned}$$

It follows that

$$\frac{\varepsilon}{80}n^2 < 2\delta n^2 + 9n < 3\delta n^2$$

for n sufficiently large; hence, $\delta > \varepsilon/240$, a contradiction. Therefore, $x < \varepsilon n/4$. Now substitute (6)–(8) and the trivial $e_r(X) < x^2$ into (5) to obtain

$$\begin{aligned}\left(\frac{1}{2}-\delta\right)n^2 < e_r(S) &< \left(\frac{1}{2}+\delta\right)n_1^2 + x^2 + \left(\frac{3}{4}+\varepsilon\right)n_1x + \frac{1}{4}n_1 \\ &\quad + 8n + yn_1 + yx \\ &= \left(\frac{1}{2}+\delta\right)(n_1+x+y)^2 + \left(\frac{1}{2}-\delta\right)x^2 - \left(\frac{1}{2}+\delta\right)y^2 \\ &\quad - 2\delta yn_1 - 2\delta yx - \left(\frac{1}{4}+2\delta-\varepsilon\right)n_1x + 8n + \frac{1}{4}n_1 \\ &< \left(\frac{1}{2}+\delta\right)n^2 + \left(\frac{1}{2}-\delta\right)x^2 - \left(\frac{1}{2}+\delta\right)y^2 + 9n,\end{aligned}$$

from which it follows that

$$\left(\frac{1}{2}+\delta\right)y^2 < 2\delta n^2 + \left(\frac{1}{2}-\delta\right)x^2 + \delta n^2 < 3\delta n^2 + \frac{1}{2}\left(\frac{\varepsilon n}{4}\right)^2$$

for n sufficiently large, and

$$y^2 < \left(\left(\frac{\varepsilon}{4}\right)^2 + 6\delta\right)n^2 < \left(\frac{\varepsilon}{3}\right)^2 n^2.$$

Thus, $y < \varepsilon n/3$, and it follows that $|X \cup Y| = x + y < \varepsilon n/4 + \varepsilon n/3$, which finishes the proof of Theorem C. □

6 Proof of Theorem B

By Theorem C, extremal favorite distance digraphs are unit distance graphs after scaling and up to an exceptional set S_0 of $o(n)$ points. Similarly, by removing a set S_0 of $o(n)$ points from an extremal furthest-neighbor digraph, we obtain a maximum distance graph. (Note that none of the furthest distances changes when restricted to the Lenz configuration $S \setminus S_0$.)

Our goal is to show that there are, in fact, no exceptional points in an extremal configuration, that is, that $r|_{S_0} \equiv 1$. We do this by some careful counting. In particular, we need to understand how quickly the functions $u_d(n)$ and $M_d(n)$ grow; that is, we need lower bounds for $u_d(n) - u_d(n-k)$ and $M_d(n) - M_d(n-k)$, where k is small. The exact values of $u_d(n)$ and $M_d(n)$ are known for all sufficiently large n depending on d, except in the case of $u_d(n)$ for odd $d \geq 5$. Thus, in these cases we may simply calculate. Where we don't know the exact values, we have to use our knowledge of the structure of extremal unit distance and diameter graphs (Theorem 3).

In the next two lemmas, we state the values for $M_d(n)$ as well as $u_4(n)$. [The exact values of $u_d(n)$ for even $d \geq 6$ can be found in [12].] Here $t_p(n)$ denotes the number of edges of a *Turán p-partite graph* on n vertices, that is, of a complete p-partite graph with the n vertices divided into p parts as equally as possible.

Lemma 6.5 (Brass [4], Van Wamelen [14]). *For all $n \geq 5$,*

$$u_4(n) = \begin{cases} t_2(n) + n & \text{if } n \text{ is divisible by } 8 \text{ or } 10, \\ t_2(n) + n - 1 & \text{otherwise.} \end{cases}$$

Lemma 6.6 ([12]). *For all sufficiently large n (depending on d),*

$$M_4(n) = \begin{cases} t_2(n) + \lceil n/2 \rceil + 1 & \text{if} \quad n \not\equiv 3 \pmod 4, \\ t_2(n) + \lceil n/2 \rceil & \text{if} \quad n \equiv 3 \pmod 4; \end{cases}$$

$$M_5(n) = t_2(n) + n;$$

$$M_d(n) = t_p(n) + p \quad \text{for even } d \geq 6, \text{ where } p = d/2;$$

$$M_d(n) = t_p(n) + \lceil n/p \rceil + p - 1 \quad \text{for odd } d \geq 7, \text{ where } p = \lfloor d/2 \rfloor.$$

Lemma 6.7. *For any $d \geq 4$, there exists $N = N_d \geq 1$ such that for any n and k such that $n > k \geq 1$ and $n-k \geq N$,*

$$u_d(n) - u_d(n-k) \geq \left(1 - \frac{1}{p}\right) k(n-k), \tag{10}$$

$$M_d(n) - M_d(n-k) \geq \left(1 - \frac{1}{p}\right) k(n-k), \tag{11}$$

$$u_5(n) - u_5(n-k) \geq \frac{1}{2}k(n-k) + \frac{k^2+2k-1}{4}, \tag{12}$$

and for $n-k \geq N_5$,

$$M_5(n) - M_5(n-k) \geq \frac{1}{2}k(n-k) + \frac{k^2+4k-1}{4}. \tag{13}$$

Also, for n sufficiently large,

$$u_4(n) - u_4(n-1) = \frac{n-1}{2} \quad \text{only if } 8 \mid n-1 \text{ or } 10 \mid n-1. \tag{14}$$

Proof. Since Lemmas 6.5 and 6.6 provide the exact values of $u_4(n)$ and $M_d(n)$, $d \geq 4$, the inequalities (11), (13) and (14) can be obtained by simple calculations. We omit the details, except to note that $t_p(n) - t_p(n-k) \geq (1-1/p)k(n-k)$, as can be seen by taking a Turán p-partite graph on $n-k$ vertices and adding k new vertices to the smallest class.

We next prove the remaining inequalities (10) and (12). Since these all involve $u_d(n)$, for which we do not have exact values when $d \geq 5$ is odd, we give a structural argument. Consider a set S of $n-k$ points in $\mathbb{R}^d$ that is extremal with respect to unit distances; that is, $u(S) = u_d(n-k)$. By Theorem 3, S is a Lenz configuration if $n-k$ is sufficiently large. In particular, S can be partitioned into $p = \lfloor d/2 \rfloor$ parts $S_1, \dots, S_p$ with each part lying on a circle (except if d is odd when S_1 lies on a sphere) such that the distance between any two points on different circles (on a circle and the sphere, respectively) equals 1. Let $i \in [p]$ be such that $|S_i| = \min\{|S_1|, \dots, |S_p|\}$. Thus, $|S_i| \leq (n-k)/p$. Choose a set T of any k points on the circle (or sphere) containing S_i disjoint from S_i. Then $S \cup T$ contains n points and has at least $k|S \setminus S_i| \geq k(1 - 1/p)(n-k)$ additional unit distance pairs. This establishes (10).

Next consider (12). Here $S \subset \Sigma_1 \cup C_2$, where Σ_1 is a 2-sphere and C_2 a circle, with any point on Σ_1 and any point on C_2 at unit distance. Now add k new points to S in the following more careful way. If k is even, add $k/2$ points to each of Σ_1 and C_2. This creates

$$\frac{k}{2}|S \cap \Sigma_1| + \frac{k}{2}|S \cap C_2| = k(n-k)/2$$

unit distance pairs from the new points to S and $k^2/4$ unit distance pairs between the new points. Since $r_1 > 1/2$ [otherwise, $u_d(n-k) = u(S) \leq t_2(n-k) + O(n)$, a contradiction], it is possible to choose each new point on Σ_1 at unit distance to some point of $S\cap\Sigma_1$. We obtain a set S' of n points with at least

$$u_5(n-k)+\frac{1}{2}k(n-k)+\frac{k^2}{4}+\frac{k}{2}$$

unit distance pairs. Therefore,

$$u_5(n)\geq u(S')\geq u_5(n-k)+\frac{1}{2}k(n-k)+\frac{k^2}{4}+\frac{k}{2},$$

which proves (12) when k is even. Now let k be odd. If we place $(k-1)/2$ points on Σ_1 and $(k+1)/2$ points on C_2, this creates as before

$$\begin{aligned}&\frac{k-1}{2}\,|S\cap C_2|+\frac{k+1}{2}\,|S\cap\Sigma_1|+\frac{k^2-1}{4}+\frac{k-1}{2}\\&\quad=\frac{1}{2}k(n-k)+\frac{1}{2}(|S\cap\Sigma_1|-|S\cap C_2|)+\frac{k^2-1}{4}+\frac{k-1}{2}\end{aligned}\tag{15}$$

additional unit distance pairs. If instead we place $(k+1)/2$ points on Σ_1 and $(k-1)/2$ points on C_2, the number of additional unit distance pairs created is

$$\begin{aligned}&\frac{k+1}{2}\,|S\cap C_2|+\frac{k-1}{2}\,|S\cap\Sigma_1|+\frac{k^2-1}{4}+\frac{k+1}{2}\\&\quad=\frac{1}{2}k(n-k)+\frac{1}{2}(|S\cap C_2|-|S\cap\Sigma_1|)+\frac{k^2-1}{4}+\frac{k+1}{2}.\end{aligned}\tag{16}$$

It is always possible to attain at least the average of (15) and (16), which equals the right-hand side of (12). □

In the above lemma it is tempting to try to prove the inequalities (11) and (13) also with the use of Theorem 3. However, we should then be careful in how we choose the k points to be added to the extremal Lenz configuration on $n-k$ points, so as not to change the furthest distances among the original $n-k$ points. Although this is possible, the case $d=5$ requires a very detailed consideration of the extremal five-dimensional diameter graphs as determined in [12]. It is much simpler to instead use the estimates from Lemma 6.6 and calculate.

We can now start with the proof of Theorem B. Let $d\geq 4$, $p=\lfloor d/2\rfloor$, and let $S\subset\mathbb{R}^d$ with $|S|=n$ and $r\colon S\to(0,\infty)$ determine an extremal favorite distance digraph [or let S determine an extremal furthest neighbor digraph respectively, and then continue to write $r(\boldsymbol{x})=D_S(\boldsymbol{x})$ for $\boldsymbol{x}\in S$].

Apply Theorem C. Thus, if n is sufficiently large depending on d, $\mathbb{R}^d$ has an orthogonal decomposition $V_1\oplus\ldots V_p$ with $\dim V_i=2$ (except when d is odd, $\dim V_1=3$) such that after scaling and translation of S, there is a Lenz system $(C_1,\ldots,C_p)$ for d even, $(\Sigma_1,C_2,\ldots,C_p)$ for d odd, and a partition $S_0,S_1,\ldots,S_p$ of S

with $|S_0| = o(n)$, $|S_i| = \frac{n}{p} + o(n)$, and $S_i \subset C_i$, where C_i is a circle with center o and radius r_i in V_i for $i \in [p]$ (except if d is odd and $i = 1$, where $S_1 \subset \Sigma_1 \subset V_1$) such that $r_i^2 + r_j^2 = 1$ for all distinct i, j. Also, $r|_{S \setminus S_0} \equiv 1$.

Let $T := \{\boldsymbol{x} \in S_0 : r(\boldsymbol{x}) \neq 1\}$. If we can show that $T = \varnothing$, then $r \equiv 1$ and S would consequently determine an extremal unit distance graph (extremal diameter graph, respectively), since $2u(S) = e_r(S) \geq 2u_d(n)$ [respectively, $2M(S) = e_r(S) \geq 2M_d(n)$]. It would then follow from Theorem 3 that S is a Lenz configuration for sufficiently large n. In the exceptional case of favorite distances in dimension 4, we show instead that if $T \neq \varnothing$, then $T = \{\boldsymbol{o}\}$. As in the proof of Theorem C, the dimensions $d \geq 6$ are disposed of very quickly, and the case $d = 5$ takes the most work.

Write $k := |T|$. Suppose that $k \neq 0$. We aim to find a contradiction except in the four-dimensional case, where we'll prove that $k = 1$ and $T = \{\boldsymbol{o}\}$, $r(\boldsymbol{o}) = r_1 = r_2 = 1/\sqrt{2}$.

We estimate as follows:

$$\begin{aligned} 2u_d(n) \leq e_r(S) &= e_r(S \setminus T) + e_r(S \setminus T, T) + e_r(T, S \setminus T) + e_r(T) \\ &\leq 2u_d(n-k) + e_r(S \setminus T, T) + e_r(T, S \setminus T) + e_r(T). \end{aligned} \tag{17}$$

This, together with (10) of Lemma 6.7, gives

$$\begin{aligned} 2\left(1 - \frac{1}{p}\right) k(n-k) &\leq 2u_d(n) - 2u_d(n-k) \\ &\leq e_r(S \setminus T, T) + e_r(T, S \setminus T) + e_r(T). \end{aligned}$$

Using instead (11) (for the case of furthest neighbors) gives the same bounds, so in all cases we have

$$2\left(1 - \frac{1}{p}\right) k(n-k) \leq e_r(S \setminus T, T) + e_r(T, S \setminus T) + e_r(T). \tag{18}$$

Since $r(\boldsymbol{x}) \neq 1$ for all $\boldsymbol{x} \in T$, $\boldsymbol{x}$ is not adjacent to any point from $S \setminus T$ in the graph $G_r^2(S)$ of double edges, so

$$e_r(S \setminus T, T) + e_r(T, S \setminus T) \leq k(n-k). \tag{19}$$

Substituting this and the trivial bound $e_r(T) \leq k(k-1)$ into (18), we obtain

$$2\left(1 - \frac{1}{p}\right) k(n-k) \leq k(n-1).$$

Since $k = o(n)$, we obtain a contradiction for sufficiently large n if $p \geq 3$. This finishes the proof for the cases $d \geq 6$.

Now assume that $d \in \{4,5\}$. Suppose that for some $x \in T$, $e_r(S_2,\boldsymbol{x}) + e_r(\boldsymbol{x},S_2) \le 4$. Then we may improve (19) to

$$e_r(S \setminus T, T) + e_r(T, S \setminus T) \le k(n-k) - |S_2| + 4.$$

Substituting this and $e_r(T) \le k(k-1)$ into (18), we obtain $k(n-k) \le k(n-1) - |S_2| + 4$; hence, $|S_2| \le 4 + k(k-1)$. This contradicts $|S_2| = n/2 + o(n)$ for n sufficiently large.

Therefore, for all $\boldsymbol{x} \in T$, we have $e_r(S_2,\boldsymbol{x}) + e_r(\boldsymbol{x},S_2) \ge 5$, which implies either $e_r(S_2,\boldsymbol{x}) \ge 3$ or $e_r(\boldsymbol{x},S_2) \ge 3$. Either case gives that $\boldsymbol{x}$ is equidistant to the circle C_2. Therefore, $x \in V_1$.

We have shown that $T \subset V_1$.

We can now finish the case $d = 4$. Symmetry gives that $T \subset V_2$ as well; hence, $T = \{\boldsymbol{o}\}$. Therefore, $\boldsymbol{o}$ must have the same distance $r(\boldsymbol{o})$ to C_1 and C_2, and it follows that $r(\boldsymbol{o}) = r_1 = r_2 = 1/\sqrt{2}$ and $e_r(S \setminus \{\boldsymbol{o}\}, \boldsymbol{o}) + e_r(\boldsymbol{o}, S \setminus \{\boldsymbol{o}\}) = n-1$. In the favorite distance case, we obtain from (17) that $u_4(n) - u_4(n-1) \le (n-1)/2$. Combined with (10) of Lemma 6.7, we obtain that equality holds; hence, $8 \mid n-1$ or $10 \mid n-1$ by (14), and $S \setminus \{\boldsymbol{o}\}$ is an extremal unit distance configuration. Inspection of the extremal configurations [4, 14] shows that when $8 \nmid n-1$ and $10 \mid n-1$, the two circles are necessarily of different radii. In our case, we must therefore have $8 \mid n-1$. Then the extremal unit distance configurations on $n-1$ points are formed by the vertices of $(n-1)/8$ unit squares inscribed in each C_i [4].

In the furthest-neighbor case, we obtain similarly as above that $M_4(n) - M_4(n-1) \le (n-1)/2$. Again, by (11), equality holds and $S \setminus \{\boldsymbol{o}\}$ is an extremal diameter configuration. However, it is easily seen that when $r_1 = r_2 = 1/\sqrt{2}$, the maximum number of diameter pairs in a set of n points in $C_1 \cup C_2$ is $t_2(n) + 2$, which contradicts Lemma 6.6 for sufficiently large n. This finishes the proof for the case $d = 4$.

Now consider the case $d = 5$. Suppose that $e_r(S_1,\boldsymbol{x}) + e_r(\boldsymbol{x},S_1) < n/3$ for some $\boldsymbol{x} \in T$. Then we may improve (19) to

$$e_r(S \setminus T, T) + e_r(T, S \setminus T) < k(n-k) - |S_1| + \frac{n}{3},$$

which, when substituted together with $e_r(T) \le k(k-1)$ into (18), gives $k(n-k) < k(n-1) - |S_1| + n/3$ and subsequently, $|S_1| < n/3 + k(k-1)$, which contradicts $|S_1| = n/2 + o(n)$ for n sufficiently large.

Therefore, for all $\boldsymbol{x} \in T$, we have $e_r(S_1,\boldsymbol{x}) + e_r(\boldsymbol{x},S_1) \ge n/3$. This will enable us to show that T lies on a straight line through the origin. Suppose then that for some two $\boldsymbol{x},\boldsymbol{x}' \in T \setminus \{\boldsymbol{o}\}$, the lines $\boldsymbol{ox}$ and $\boldsymbol{ox}'$ are not parallel. Then at least $n/3$ points of S_1 lie on two circles that are both normal to $\boldsymbol{ox}$, and similarly, at least $n/3$ points of S_1 lie on two circles normal to $\boldsymbol{ox}'$. Since the intersection of these two unions of circles contains at most 8 points, we obtain

$$n/2 + o(n) = |S_1| \ge 2 \cdot \frac{n}{3} - 8,$$

a contradiction for sufficiently large n.

It follows that T lies on a line ℓ, say, through the origin. Since there are at most 2 points on ℓ at distance $r(\boldsymbol{x})$ to $\boldsymbol{x}$, it follows that $e_r(\boldsymbol{x},T) \leq 2$ for all $\boldsymbol{x} \in T$, and when $\boldsymbol{x}$ is the first or last point of T on ℓ, $e_r(\boldsymbol{x},T) \leq 1$. It follows that $e_r(T) \leq 2k-2$ (keeping in mind that $T \neq \varnothing$ by assumption).

In the case of extremal favorite distance digraphs, bounds (17), (19) and $e_r(T) \leq 2k-2$, together with (12) of Lemma 6.7, give

$$k(n-k)+\frac{k^2+2k-1}{2} \leq 2u_5(n)-2u_5(n-k) \leq k(n-k)+2k-2,$$

which simplifies to $(k-1)^2+2 \leq 0$, a contradiction.

For extremal furthest-neighbor digraphs, a similar calculation (now using (13) instead of (12)) gives that

$$k(n-k)+\frac{k^2+4k-1}{2} \leq 2M_5(n)-2M_5(n-k) \leq k(n-k)+2k-2,$$

which simplifies to $k^2+3 \leq 0$, another contradiction.

We have shown that $k=0$ in all cases when $d=5$, and it follows that S is a Lenz construction. □

Acknowledgements The author would like to thank the anonymous referee for careful proofreading and good advice on a previous version.

References

1. B. Aronov, M. Sharir, Cutting circles into pseudo-segments and improved bounds for incidences. Discrete Comput. Geom. **28**, 475–49 (2002)
2. D. Avis, The number of furthest neighbour pairs of a finite planar set. Am. Math. Mon. **91**, 417–420 (1984)
3. D. Avis, P, Erdős, J. Pach, Repeated distances in space. Graphs Combin. **44**, 207–217 (1988)
4. P. Brass, On the maximum number of unit distances among n points in dimension four, in *Intuitive Geometry*, ed. by I. Bárány et al. Bolyai Soc. Mathematical Studies, vol. 6 (1997), pp. 277–290. See also the review of this paper in Mathematical Reviews MR 98j:52030.
5. P. Brass, W.O.J. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, Berlin, 2005)
6. W.G. Brown, M. Simonovits, Extremal multigraph and digraph problems, in *Paul Erdős and His Mathematics*, vol. 2, ed. by G. Halász et al. Bolyai Soc. Mathematical Studies, vol. 11 (2002), pp. 157–203
7. G. Csizmadia, Furthest neighbours in space. Discrete Math. **150**, 81–88 (1996)
8. H. Edelsbrunner, S. S. Skiena, On the number of furthest neighbour pairs in a point set. Am. Math. Mon. **96**, 614–618 (1989)
9. P. Erdős, On sets of distances of n points in Euclidean space. Magyar Tud. Akad. Mat. Kut. Int. Közl. **5**, 165–169 (1960)
10. P. Erdős, On some applications of graph theory to geometry. Can. J. Math. **19**, 968–971 (1967)
11. P. Erdős, J. Pach, Variations on the theme of repeated distances. Combinatorica **10**, 261–269 (1990)

12. K.J. Swanepoel, Unit distances and diameters in Euclidean spaces. Discrete Comput. Geom. **41**, 1–27 (2009)
13. K.J. Swanepoel, *Favourite distances in* 3-*space*, Manuscript.
14. P. van Wamelen, The maximum number of unit distances among *n* points in dimension four. Beiträge Algebra Geom. **40**, 475–477 (1999)

Intersection Patterns of Convex Sets via Simplicial Complexes: A Survey

Martin Tancer

Abstract The task of this survey is to present various results on intersection patterns of convex sets. One of the main tools for studying intersection patterns is a point of view via simplicial complexes. We recall the definitions of so-called d-representable, d-collapsible, and d-Leray simplicial complexes, which are very useful for this study. We study the differences among these notions and also focus on computational complexity for recognizing them. A list of Helly-type theorems is presented in the survey. We also discuss the important role played by the above-mentioned notions for the theorems. We also consider intersection patterns of good covers, which generalize collections of convex sets (the sets may be "curvy"; however, their intersections cannot be too complicated). We mainly focus on new results.

Mathematics Subject Classification (2010): primary 52A35, secondary 05E45, 52A20

1 Introduction

An important branch of combinatorial geometry regards studying intersection patterns of convex sets. Research in this area was initiated by a theorem of Helly [15], which can be formulated as follows: If $C_1, \ldots, C_n$ are convex sets in $\mathbb{R}^d$, $n \geq d+1$, and any collection of $d+1$ sets among $C_1, \ldots, C_n$ has a nonempty

M. Tancer (✉)
Faculty of Mathematics and Physics, Department of Applied Mathematics and Institute for Theoretical Computer Science, Charles University, Malostranské nám. 25, 118 00 Prague, Czech Republic
e-mail: tancer@kam.mff.cuni.cz

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_28,

intersection, then all the sets have a common point. We will focus on results of a similar spirit; however, we have to set up some notations first.

1.1 Simplicial Complexes

First, we recall simplicial complexes, which provide a convenient language for studying intersection patterns of convex sets. We assume that the reader is familiar with simplicial complexes, and thus we only briefly mention the basics. For further details, the reader is referred to books such as [14, 30, 32].

We deal with finite *abstract simplicial complexes*, i.e., collections K of subsets of a finite set X such that if $\alpha \in \mathsf{K}$ and $\beta \subset \alpha$, then $\beta \in \mathsf{K}$. Elements of K are *faces* of K. The *dimension* of a face $\alpha \in \mathsf{K}$ is defined as $|\alpha| - 1$; i-dimensional faces for $i \in \{0,1,2\}$ are *vertices*, *edges*, and *triangles*, respectively. The *dimension* of a simplicial complex is the maximum dimension of its faces. Graphs coincide with one-dimensional simplicial complexes. If V' is a subset of vertices of K, then the *induced subcomplex* $\mathsf{K}[V']$ is a complex of faces $\alpha \in \mathsf{K}$ such that $\alpha \subseteq V'$. We use the notation $\mathsf{L} \leq \mathsf{K}$ to point out that L is an induced subcomplex of K. An *m-skeleton* of K is a simplicial complex consisting of faces of K of dimension at most m. We denote it by $\mathsf{K}^{(m)}$. The m-dimensional *full simplex*, Δ_m, is a simplicial complex with the vertex set $\{1,\dots,m+1\}$ and all possible faces. Let Y be a set of affinely independent points in $\mathbb{R}^{|X|-1}$ such that there is a bijection $f\colon X \to Y$. The *geometric realization* of K, denoted by $|\mathsf{K}|$, is the topological space $\bigcup\{\operatorname{conv} f(\alpha)\colon \alpha \in \mathsf{K}\}$, where conv denotes the convex hull.

1.2 d-Representable Complexes

Let $\mathcal{C}$ be a collection of some subsets of a given set X. The *nerve* of $\mathcal{C}$, denoted by $\mathsf{N}(\mathcal{C})$, is a simplicial complex whose vertices are the sets in $\mathcal{C}$ and whose faces are subcollections $\{C_1,\dots,C_k\} \subseteq \mathcal{C}$ such that the intersection $C_1 \cap \dots \cap C_k$ is nonempty. The notion of a nerve is designed to record the "intersection pattern" of the sets in $\mathcal{C}$.

A simplicial complex K is *d-representable* if it is isomorphic to the nerve of a finite collection of convex sets in $\mathbb{R}^d$. Such a collection of convex sets is called a *d-representation* for K. d-representable simplicial complexes are the central objects of our study in this survey. As mentioned above, they precisely record all possible intersection patterns of finite collections of convex sets in $\mathbb{R}^d$.

Using the notion of d-representability, the Helly theorem can be reformulated as follows: If a d-representable complex on at least $d+1$ vertices contains all possible d-faces, then it is already a full simplex. This statement is, via induction, equivalent with the following: A d-representable simplicial complex does not contain an induced k-dimensional *simplicial hole* for $k \geq d$, i.e., a complex isomorphic to $\Delta_{k+1}^{(k)}$. (Note that "hole" refers here to a hole in a certain topological space, and in particular

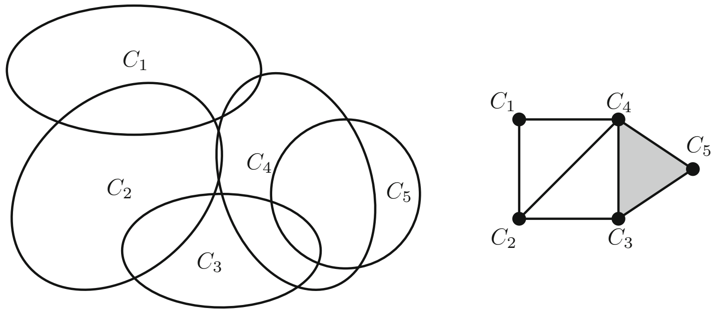


Fig. 1 A 2-representable complex and its nerve

this notion is different from a k-hole in the context of Horton sets. The dimension of the hole refers to the dimension of the boundary rather than to the dimension of the missing part. We also remark that a geometric representation of a k-dimensional simplicial hole is homeomorphic to the *k-sphere* S^k; and it is the simplest way to obtain the k-sphere as a simplicial complex.) The *Helly number* of a simplicial complex K is the total number of vertices of the largest simplicial hole in K (i.e., the dimension of the hole plus 2). Another reformulation of the Helly theorem thus states that the Helly number of a d-representable complex is at most $d+1$.

Example 1.1. Figure 1 shows a collection $\mathcal{C} = \{C_1, C_2, C_3, C_4, C_5\}$ of convex sets (on the left) and their nerve (on the right). In other words, the simplicial complex on the right is 2-representable and $\mathcal{C}$ is a 2-representation of it. The Helly number of this collection equals 3.

1.3 What Is in the Survey?

The task of the survey is to give an overview of recent developments in the study of intersection patterns of finite collections of convex sets. We are mainly focusing on the description of intersection patterns via simplicial complexes. Apart from geometrical "ad hoc" arguments, there are two other important approaches we are going to discuss. The first is combinatorial and regards d-collapsibility. The second one is topological and regards the Leray number of a simplicial complex. These two approaches are mainly discussed in the following two sections. Algorithmic aspects of the recognition of intersection patterns are briefly discussed in Sect. 4. In Sect. 5, we mention some properties of good covers as a natural generalization of collections of convex sets. Finally, Sect. 6 contains a list of theorems on intersection patterns. This list appears at the end of the chapter so that all the necessary terminology is already built up. However, many of the results in Sect. 6 can be understood without a detailed study of the previous sections.

Since our task is to cover only a very selected part of convex geometry, we refer the reader to other sources regarding the related areas. In particular, we refer to [29] and the references therein for an extended basic overview on *discrete convex geometry*, including Radon-, Carathéodory-, and Tverberg-type theorems; we also refer to [9] for another point of view on the area; and to [12] for results on transversals to convex sets. We do not focus on the theory of f-vectors. For readers interested in f-vectors, we refer to [6] or to [18] for a useful method for investigating f-vectors (and related also to other branches mentioned here).

2 *d*-Collapsible and *d*-Leray Complexes

There are two other important classes of simplicial complexes related to the d-representable ones. Informally, a simplicial complex is d-collapsible if it can be vanished by removing faces of dimension at most $d-1$ that are contained in a single maximal face; a simplicial complex is d-Leray if its induced subcomplexes do not contain, homologically, holes of dimension d or more.

Wegner [41] proved that d-representable simplicial complexes are d-collapsible and also that d-collapsible complexes are d-Leray.

Now, we will precisely define d-collapsible complexes and then d-Leray complexes.

Let K be a simplicial complex. Let T be the collection of inclusion-wise maximal faces of K. A face σ is *d-collapsible* if there is only one face $\tau \in T$ containing σ (possibly $\sigma = \tau$), and moreover $\dim \sigma \leq d-1$. The simplicial complex

$$\mathsf{K}' := \mathsf{K} \setminus \{\eta \in \mathsf{K} : \eta \supseteq \sigma\}$$

is an *elementary d-collapse* of K. For such a situation, we use the notation $\mathsf{K} \to \mathsf{K}'$. A simplicial complex is *d-collapsible* if there is a sequence,

$$\mathsf{K} \to \mathsf{K}_1 \to \mathsf{K}_2 \to \cdots \to \emptyset,$$

of elementary d-collapses ending with an empty complex.

Example 2.1. A simplicial complex K consisting of a full tetrahedron, two full triangles, and one hollow triangle in Fig. 2 is 2-collapsible. For a proof, there is a 2-collapsing of K drawn on the picture. In every step, the faces σ and τ are indicated.

A simplicial K complex is *d-Leray* if the ith reduced homology group $\tilde{H}_i(\mathsf{L})$ (over $\mathbb{Q}$) vanishes for every induced subcomplex $\mathsf{L} \leq \mathsf{K}$ and every $i \geq d$.

We mention several remarks regarding d-collapsible and d-Leray complexes. Deeper properties of these complexes are studied in the following sections.

- The fact that d-collapsible complexes are d-Leray is simple (for a reader familiar with homology) since d-collapsing does not affect the homology of dimension d or more.

  It is a bit less trivial to show that a d-representable complex K is d-collapsible. The idea is to slide a generic hyperplane (from infinity to minus infinity) over a

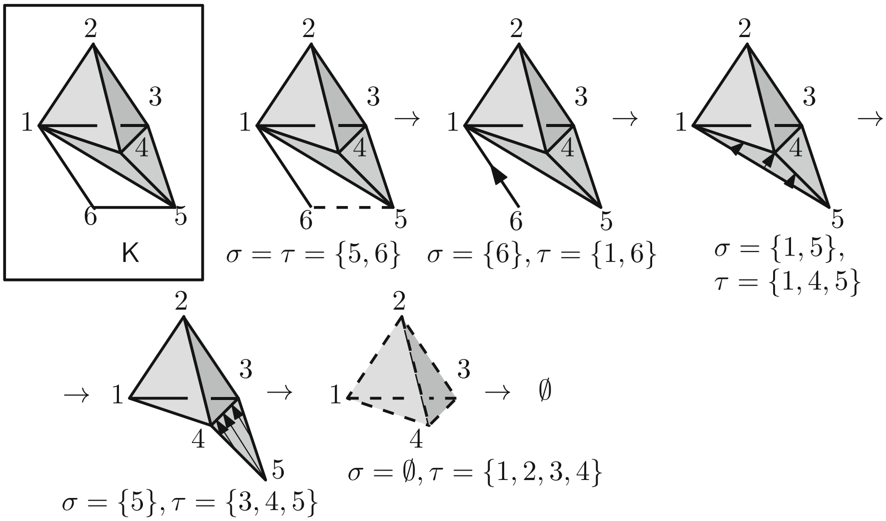


Fig. 2 A 2-collapsing of a simplicial complex

d-representation for K and gradually cut off whatever is on the positive side of the hyperplane. See Fig. 3 and the text bellow the picture for a more detailed sketch. The reader is referred to [41] for full details.

The inclusion of d-representable complexes in d-Leray complexes can be also deduced, without using Wegner's results, from the nerve theorem (see Theorem 5.1).

- A d-dimensional simplicial complex is $(d+1)$-collapsible and hence also $(d+1)$-Leray. For a complex K, the smallest possible ℓ such that K is d-Leray is traditionally called the *Leray number* of K.
- Neither d-representability, d-collapsibility nor the Leray number is an invariant under a homeomorphism: The full simplex Δ_m is 0-representable; however, its barycentric subdivision is not even $(m-1)$-Leray, since it contains an $(m-1)$-sphere as an induced subcomplex.
- It is not very difficult to see that every induced subcomplex of a d-collapsible complex is again d-collapsible. If $\mathsf{K}[V'] \leq \mathsf{K}$ and $\mathsf{K} \to \mathsf{K}_1 \to \cdots \to \emptyset$ is a d-collapsing of K, then $\mathsf{K}[V'] \to \mathsf{K}_1[V'] \to \cdots \to \emptyset = \emptyset[V']$ is a d-collapsing for $\mathsf{K}[V']$, where some steps are possibly trivial; i.e., $\mathsf{K}_i[V'] = \mathsf{K}_{i+1}[V']$.
- The Helly theorem easily follows from the fact that d-representable complexes are contained in d-collapsible ones (or d-Leray ones): We have that a d-dimensional simplicial hole is neither d-collapsible (nor d-Leray).

  On the other hand, these two notions provide (much) stronger limitations to intersection patterns than the Helly theorem. For instance, they also exclude (in dimension 2) the boundary of the octahedron (i.e., the simplicial complex with vertices $\{-3,-2,-1,1,2,3\}$ and faces α such that there is no $i \in \{1,2,3\}$ with $-i, i \in \alpha$) or a triangulation of a torus.

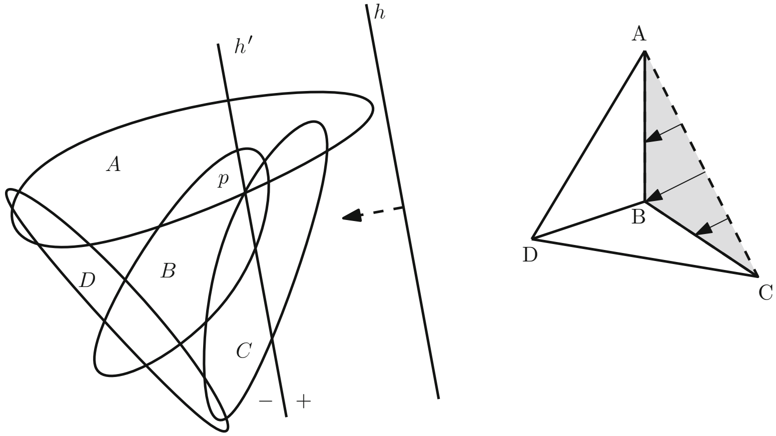


Fig. 3 A schematic sketch of the proof of Wegner's theorem. A generic hyperplane h is slid from infinity to minus infinity until there is a nontrivial intersection of the convex sets on its positive side. In this case, it slides to h' and cuts off $A \cap B \cap C$ (it also cuts off $A \cap C$, but for the moment, we consider a maximal collection). From genericity, there is a single point $p \in A \cap B \cap C \cap h'$. It can be shown (using Helly's theorem) that there are only at most d sets of the starting collection necessary to obtain p. In this case, $\{p\} = A \cap C \cap h'$. Thus, we obtain a d-collapse with $\sigma = \{A, C\}$ and $\tau = \{A, B, C\}$. Finally, $A \cap (h')^-, \dots, D \cap (h')^-$ form a d-representation for the resulting collapsed complex; thus, the procedure can be repeated

The gaps among these notions are discussed in more detail in the following section.

3 Gaps Among the Notions

In this section, we provide an overview of how the notions of d-representable, d-collapsible, and d-Leray complexes differ. We also relate these differences with the dimension of the complex.

3.1 Every Finite Simplicial Complex Is d-Representable for d Big Enough

Let K be a simplicial complex on vertex set $\{1, \dots, n\}$. Let $x_1, \dots, x_n$ be affinely independent points in $\mathbb{R}^{n-1}$ (i.e., they form a simplex). For a nonempty face $\alpha = \{a_1, \dots, a_t\} \in \mathsf{K}$, let b_α be the barycenter of the points $x_{a_1}, \dots, x_{a_t}$. Then for $i \in \{1, \dots, n\}$, we set $C_i := \operatorname{conv}\{b_\alpha : i \in \alpha, \alpha \in \mathsf{K}\}$. The reader is welcome to check that sets $C_{i_1}, \dots, C_{i_k}$ intersect if and only if $\{i_1, \dots, i_k\} \in \mathsf{K}$. Thus, the nerve of $C_1, \dots, C_n$

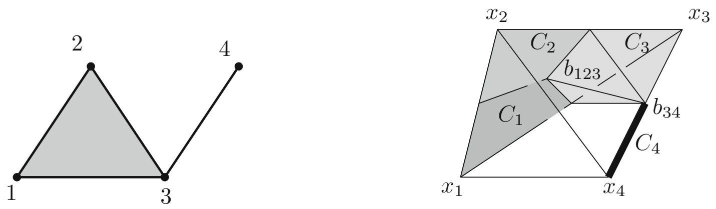


Fig. 4 Representing a complex

is isomorphic to K. See Fig. 4 for an illustration. (If we really did not care about the dimension, it would be even easier to check the situation where the points b_α are set to be the vertices of a simplex of dimension $|\mathsf{K}|-1$.)

There is, however, another way to obtain a representation of a complex depending on the dimension of the complex.

Theorem 3.1 (Wegner [40], Perel'man [34]). *Let* K *be a d-dimensional simplicial complex. Then* K *is* $(2d+1)$*-representable.*

The value $2d+1$ in Theorem 3.1 is the lowest possible. For example, the barycentric subdivisions of the d-skeleton of a $(2d+2)$-dimensional simplex is not $2d$-representable. Case $d=1$ was established by Wegner [40]; the general case is proved in [38].

The references for Theorem 3.1 are due to Eckhoff [9]. (Perel'man rediscovered Wegner's result.) Unfortunately, I have not been able to check these sources in detail (the first one is in German, the second one is in Russian). Thus, I instead supply an idea of a proof (communicated by Jiří Matoušek).

Sketch of a proof of Theorem 3.1. Let K be a d-representable complex with n vertices.

A *k-neighborly* polytope is a convex polytope such that every k vertices form a face of the polytope. It is well known that there are $2k$-dimensional k-neighborly polytopes with an arbitrary number of vertices for every $k \geq 1$. For instance, *cyclic polytopes* satisfy this property. (See, e.g., [29] for a background on convex polytopes including cyclic polytopes.)

Let Q be a $(2d+2)$-dimensional $(d+1)$-neighborly polytope with n vertices. Let Q^* be a polytope dual to Q. It has n facets, and any $d+1$ of its facets share a face of the polytope. Finally, we consider the *Schlegel diagram* of Q^*. The Schlegel diagram of an m-dimensional convex polytope is a projection of the polytope to $(m-1)$-space through a point beyond one of its facets (the point is very close to the facet). In particular, the facets of Q^* project to convex sets $C_1,\dots,C_n$ in $\mathbb{R}^{2d+1}$ such that each $d+1$ of them share the projection of a face of Q^* (on their boundary). Thus, if we look at the nerve N of $C_1,\dots,C_n$, then it contains a full d-skeleton of a simplex with n vertices. Therefore, without loss of generality, we can assume that K is a subcomplex of N. Let $\vartheta = \{C_{i_1},\dots,C_{i_j}\}$ be a face of N that does not belong to K.

The sets $C_{i_1}, \dots, C_{i_j}$ intersect on their boundaries, and it is possible to remove their intersection by removing a small neighborhood of $C_{i_1} \cap \dots \cap C_{i_j}$ in each of the sets while keeping the sets convex. Hence, only ϑ and the superfaces of ϑ disappear from the nerve during this procedure. After repeating the procedure, we obtain a collection of convex sets with the nerve K. □

3.2 The Gap Between Representability and Collapsibility

For $d = 0$, all three notions, 0-representable, 0-collapsible, and 0-Leray, coincide and they can be replaced with "being a simplex."

For $d = 1$, 1-representable complexes are clique complexes over interval graphs; 1-collapsible and 1-Leray complexes are clique complexes over chordal graphs (we remark that results in [26, 41] easily imply these statements).

For $d \geq 2$, there is perhaps no simple characterization of d-representable, d-collapsible, and d-Leray complexes. Wegner [41] gave an example of a complex that is 2-collapsible but not 2-representable. Matoušek and this author [31] found d-collapsible complexes that are not $(2d-2)$-representable. Later, the author [36] improved this result by finding 2-collapsible complexes that are not d-representable (for any fixed d). We present some steps of both constructions, since even the weaker construction contains some steps of their own interest.

Let E be a $(d-1)$-dimensional simplicial complex that is not embeddable in $\mathbb{R}^{2d-2}$. Such a complex already exists, for example, the *van Kampen complex* $\Delta_{2d}^{(d-1)}$; see [39], or the *Flores complex* [11], which is the join of d copies of a set of three independent points. The first example is the nerve $\mathsf{N}(\mathsf{E})$. It is d-collapsible but not $(2d-2)$-representable due to the following two propositions [31].

Proposition 3.2. *Let K be a simplicial complex such that the nerve $\mathsf{N}(\mathsf{K})$ is n-representable. Then K embeds in $\mathbb{R}^n$, even linearly.*

Proposition 3.3. *Let $\mathcal{F}$ be a family of sets, each of size at most n. Then the nerve $\mathsf{N}(\mathcal{F})$ is n-collapsible.*

The second (stronger) example regards finite projective planes seen as simplicial complexes. Let $(P, \mathcal{L})$ be a finite projective plane, where P is the set of its points and $\mathcal{L}$ is the set of its lines. There is a natural simplicial complex P associated with the projective plane. Its ground set is P and faces are the collections of points lying on a common line.

It is not hard to show that P is 2-collapsible. The nonrepresentability of P is summarized in the following theorem [36].

Theorem 3.4. *For every $d \in \mathbb{N}$, there is a $q_0 = q_0(d)$ such that if a complex P corresponds to a projective plane of order $q \geq q_0$, then P is not d-representable.*

We also sketch a proof of Theorem 3.4. The reader is referred to the original paper for full details. In contrast to the original paper, we present it in an "inequality form." For this, we need a few preliminaries.

One important ingredient is a selection theorem by Pach [33].

Theorem 3.5. *For every positive integer d, there is a constant $c = c(d) > 0$ with the following property. Let $X \subset \mathbb{R}^d$ be a finite set of points in general position. Then there are a point $a \in \mathbb{R}^d$ and disjoint subsets $Z_1, \dots, Z_{d+1}$ of X, with $|Z_i| \geq c|X|$ such that the convex hull of every transversal of $(Z_1, \dots, Z_{d+1})$ contains a.*

We recall that a *transversal* of a group of sets $(Z_1, \dots, Z_{d+1})$ is a set $\{z_1, \dots, z_{d+1}\}$ such that $z_i \in Z_i$.

Now we let $\mathcal{S}$ to be a subset of a simplicial complex K with a vertex set V where $V = \{1, \dots, n\}$ and $|\mathcal{S}| = s$. For $\mathcal{S}' \subset \mathcal{S}$, we define the *deficiency* of $\mathcal{S}'$ as the number of vertices of K that are not contained in any element of $\mathcal{S}'$, i.e., the value $n - |\bigcup \mathcal{S}'|$. A value $\rho(k)$ of a function $\rho\colon \{1, \dots, n\} \to \mathbb{N}$ is defined as the maximum number of deficiencies of sets $\mathcal{S}' \subseteq \mathcal{S}$ with k elements. Then we have the following inequality.

Proposition 3.6. *If K is d-representable, then*

$$n - (d+1)\rho(c(d)s) \leq \dim K + 1,$$

where $c(d)$ is Pach's constant.

Sketch of a proof. Let $C_1, \dots, C_n$ be the sets forming the d-representation of K, where set C_i corresponds to a vertex i. It can be assumed that these sets are open. For every $\sigma \in \mathcal{S}$, there is a point x_σ in the intersection of all C_i such that i is a vertex of σ. Let $X = \{x_\sigma : \sigma \in \mathcal{S}\}$. It can be assumed that X is in general position due to the openness of the sets C_i. So we have $Z_1, \dots, Z_{d+1} \subseteq X$ and $a \in \mathbb{R}^d$ from Theorem 3.5. For a fixed $j \in \{1, \dots, d+1\}$, the definition of ρ implies that only $\rho(c(d)s)$ sets among $C_1, \dots, C_n$ can avoid the points of the set Z_j. Thus, there is at least $n - (d+1)\rho(c(d)s)$ of the C_i that meet all Z_j, and therefore they contain a. Hence, the vertices of K corresponding to these C_i form a face of K of dimension $n - (d+1)\rho(c(d)s) - 1$. □

Theorem 3.4 follows from Proposition 3.6 when $\mathcal{S}$ is a set of all maximal simplices of a projective plane P. Then $n = s = q^2 + q + 1$, $\dim P = q + 1$, and $\rho(k) \leq (q^2 + q + 1)^{3/2}/k$ by a theorem of Alon [3, 4].

3.3 The Gap Between Collapsibility and Leray Number

Wegner showed an example of complex that is 2-Leray but not 2-collapsible, namely, a triangulation D of the dunce hat. If we consider the multiple join $\mathsf{D} \star \cdots \star \mathsf{D}$ of d copies of D, we obtain a complex that is $2d$-Leray but not $(3d-1)$-collapsible. See [31] for more details.

4 Algorithmic Perspective

As we consider different criteria for d-representability, it is also natural to ask whether there is an algorithm for recognition of d-representable complexes. We denote this algorithmic question as d-REPRESENTABILITY. More precisely, the input of this question is a simplicial complex. The size of the input is the number of faces of the complex. The value d is considered a fixed integer. The output of the algorithm is the answer whether the complex is d-representable or not.

We can also ask similar questions for d-collapsible and d-Leray complexes as relaxations of the previous problem. Thus, we have algorithmic problems d-COLLAPSIBILITY and d-LERAYNUMBER.

4.1 Representability

The first mentioned problem d-REPRESENTABILITY is perhaps the most difficult among the three algorithmic questions. It is NP-hard for $d \geq 2$. Reduction can be done in a very similar fashion as performing a reduction for hardness of recognition intersection graphs of segments [24, 25]. Full details can be found in [35]. On the other hand, it is not hard to see that there is a PSPACE algorithm for d-REPRESENTABILITY. It is based on solving systems of polynomial inequalities. See [25, Theorem 1.1(i)(a)] for a very similar reduction.

4.2 Collapsibility

It is shown in [35] that d-COLLAPSIBILITY is NP-complete for $d \geq 4$ and it is polynomial-time solvable for $d \leq 2$. For $d = 3$, the problem remains open.

4.3 Leray Number

The last question, d-LERAYNUMBER, is polynomial-time solvable. An equivalent characterization of d-Leray complexes is when induced subcomplexes are replaced with *links* of faces (including an empty face). See [20, Proposition 3.1] for a proof. The tests on links can be done in polynomial time since it is sufficient to test the homology up to the dimension of the complex.

4.4 Greedy Collapsibility

The algorithmic results above suggest that it is easier to test/compute the Leray number than collapsibility. However, if we are interested in them because of a hint for representability, computing collapsibility still can be more convenient, since d-collapsibility is closer to d-representability than the Leray number. An example from Sect. 3 is perhaps not so convincing; however, there is a more important example. As shown in Sect. 5, d-collapsibility can distinguish collections of convex sets and good covers.

An useful tool for computation could be greedy d-collapsibility. We say that a simplicial complex K is *greedily d-collapsible* if it is d-collapsible and any sequence of d-collapses of K ends up in a complex that is still d-collapsible. In other words, greedy collapsibility allows us to collapse the faces of K in whatever order without risk of a bad choice. Thus, if a complex is greedily d-collapsible, then there is a simple (greedy) algorithm for showing that it is d-collapsible. Not all d-collapsible complexes are greedily d-collapsible. Complexes that are not greedily d-collapsible for $d \geq 3$ are constructed in [35]. However, none of these complexes is d-representable. In summary, there is a hope for obtaining a simple algorithm for showing that a complex is either d-collapsible or not d-representable if the answer to the following question is true.

Problem 10. Is it true that every d-representable simplicial complex is greedily d-collapsible?

5 Good Covers

A *good cover* in $\mathbb{R}^d$ is a collection of open sets in $\mathbb{R}^d$ such that the intersection of any subcollection is either empty or contractible (in particular, the sets in the collection are contractible).[1] We consider only finite good covers. A simplicial complex is *topologically d-representable* if it is isomorphic to the nerve of a (finite) good cover in $\mathbb{R}^d$. We should emphasize that (for our purposes) a good cover need not cover whole $\mathbb{R}^d$.

Topologically d-representable complexes generalize d-representable complexes since every collection of convex sets is a good cover.

---

[1]The definition of a good cover is not fully standard in the literature. For example, it may be assumed that sets in the collection are closed instead of open, or that the intersections are homeomorphic to (open) balls instead of contractible. These differences are not essential for most of the purposes mentioned here, because all these options satisfy the assumptions of a nerve theorem (see upcoming text).

5.1 Nerve Theorems

Suppose we are given a collection $\mathcal{F}$ of subsets of $\mathbb{R}^d$. If the sets are "sufficiently nice" and also all their intersections are sufficiently nice, then the nerve of the collection, $\mathsf{N}(\mathcal{F})$, is homotopy equivalent to the union of the sets in the collection, $\bigcup \mathcal{F}$. For a weaker assumption on "sufficiently nice," it is possible to derive not necessarily homotopy equivalence, but at least equivalence on homology (up to some level). Such results are known as *homotopic/homological* nerve theorems.

We mention here one possible version (suitable for our purposes); see [14, Corollary 4G.3].

Theorem 5.1 (A homotopy nerve theorem). *Let $\mathcal{F}$ be a collection of open contractible sets in a paracompact space X such that $\bigcup \mathcal{F} = X$ and every nonempty intersection of finitely many sets in $\mathcal{F}$ is contractible (or empty). Then the nerve $\mathsf{N}(\mathcal{F})$ and X are homotopy equivalent.*

Corollary 5.2. *The nerve of every good cover is d-Leray.*

5.2 Good Covers Versus Collections of Convex Sets

Good covers have many similar properties as collections of convex sets. Many results on intersection patterns of convex sets can be generalized for good covers. We will discuss these generalizations in the following section. An exceptional case is Theorem 3.4, which cannot be generalized for good covers.

On the other hand, it is not hard to see that topologically d-representable complexes are strictly more general than d-representable complexes for $d \geq 2$. There is a less trivial example in Fig. 5 showing that there is a complex that is topologically d-representable but not d-collapsible. Originally, Wegner conjectured that there is no such example. See [37] for more details.

In addition, this example also distinguishes good covers in $\mathbb{R}^2$ and arrangements of convex sets in a topological plane:[2] An *arrangement of pseudolines* is a set of curves (called *pseudolines*) in the plane such that every two pseudolines intersect in exactly one point. The most convenient way is perhaps to think of the plane as a subset of the real projective plane; then we can allow even "parallel" pseudolines. Such an arrangement can be extended to a *topological plane*, where there is a pseudoline through every pair of points. See, e.g., [13] for more precise definitions and another background. A *convex set in a topological plane* is such a subset that every two points are connected with a "segment" of a pseudoline. The nerve of a bounded collection of convex sets in a topological plane is 2-collapsible. This can

[2] This observation is by Xavier Goaoc.

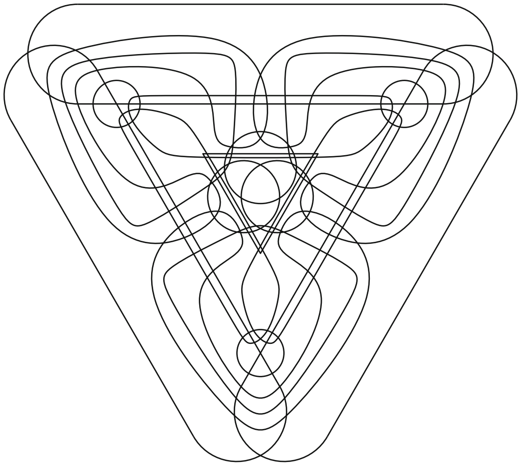


Fig. 5 A good cover $\mathcal{F}$ such that the nerve of $\mathcal{F}$ is even not d-collapsible

be shown in a very similar way as we did for Wegner's theorem on the inclusion of d-representable complexes in d-collapsible ones. Thus, the example from Fig. 5 cannot be a collection of convex sets in a topological plane.

6 Helly-Type Theorems

In this section, we give an overview of some Helly-type results on intersection patterns of convex sets. We always start with a geometrical formulation. Then we reformulate such a result via d-representable simplicial complexes. We also discuss possible extensions to d-collapsible or d-Leray complexes. (The former ones then have geometric consequences for good covers.)

6.1 The Helly Theorem

For completeness of this section, we also recall the Helly theorem mentioned in the Introduction.

We have the following geometric formulation.

Theorem 6.1 (Helly, [15]). *If $C_1, \dots, C_n$ are convex sets in $\mathbb{R}^d$, $n \geq d+1$, and any collection of $d+1$ sets among $C_1, \dots, C_n$ has a nonempty intersection, then all the sets have a common point.*

A topological extension of the Helly theorem was proved a few years later by Helly himself [16]. His setting was for good covers. We present here a setting for d-Leray complexes. Note that the theorem stated here is trivial; however, the fact that it is meaningful relies on Corollary 5.2.

Theorem 6.2. *The Helly number of a d-Leray simplicial complex is at most $d+1$.*

Theorem 6.1 is a consequence of Theorem 6.2 if it is used for d-representable complexes.

6.2 The Colorful Helly Theorem

The colorful Helly theorem regards the situation where convex sets are colored. If there are enough color classes and every rainbow collection of the colored sets contains a point in common, then the sets of a certain color class contain a point in common.

Theorem 6.3 (Colorful Helly, Lovász [27]). *Let $\mathcal{F}_1, \ldots, \mathcal{F}_{d+1}$ be families of convex sets in $\mathbb{R}^d$. Suppose that for every choice $F_1 \in \mathcal{F}_1, \ldots, F_{d+1} \in \mathcal{F}_{d+1}$, the intersection $F_1 \cap \cdots \cap F_{d+1}$ is nonempty. Then there is $i \in [d+1]$ such that the intersection of the sets in $\mathcal{F}_i$ is nonempty.*

The Helly theorem is the consequence of the colorful Helly theorem if we set $\mathcal{F}_1 = \mathcal{F}_2 = \cdots = \mathcal{F}_{d+1}$.

The reformulation via simplicial complexes is the following.

Let V be a finite set partitioned into disjoint color classes $V_1, \ldots, V_k$. A subset $W \subseteq V$ is *rainbow* if $|V_i \cap W| \leq 1$ for $i \in [k]$.

Theorem 6.4. *Let K be a d-representable simplicial complex with vertices partitioned into $d+1$ color classes. Assume that every rainbow subset of vertices is a simplex of K. Then there is a color class such that its vertices form a simplex of K.*

Let M' be a simplicial complex with the vertex set V from above whose faces are the rainbow subsets of V. It is not hard to see that M' is a matroidal complex of rank k.

Kalai and Meshulam [19] obtained the following matroidal extension of the colorful Helly theorem.

Theorem 6.5. *Let K be a d-collapsible simplicial complex on V, and let M be a matroidal complex on V with rank function ρ such that $\mathsf{M} \subseteq \mathsf{K}$. Then there is a simplex $\alpha \in \mathsf{K}$ such that $\rho(\alpha) = \rho(\mathsf{M})$ and $\rho(V \setminus \alpha) \leq d$.*

Theorem 6.5 indeed generalizes Theorem 6.4. If we set $\mathsf{M} = \mathsf{M}'$, then α contains all vertices of a certain color class (since $\rho(V \setminus \alpha) \leq d$), and moreover α even contains a vertex of every color class [since $\rho(\alpha) = \rho(\mathsf{M})$].

More importantly, Kalai and Meshulam [19] obtained a topological generalization (with a weaker conclusion, but still more general than that of Theorem 6.4).

Theorem 6.6. *Let K be a d-Leray simplicial complex on V, and let M be a matroidal complex on V with rank function ρ such that $\mathsf{M} \subseteq \mathsf{K}$. Then there is a simplex $\alpha \in \mathsf{K}$ such that $\rho(V \setminus \alpha) \leq d$.*

6.3 The Fractional Helly Theorem

Let $\mathcal{C}$ be again a collection of convex sets in $\mathbb{R}^d$ (containing at least $d+1$ sets). The Helly theorem assumes that if *every* $d+1$-tuple has a nonempty intersection, then *all* the sets have a point in common. The fractional Helly theorem is designed for a situation when *many* $d+1$-tuples have a nonempty intersection, concluding that *many* sets of the collection have a point in common.

Theorem 6.7 (Fractional Helly, Katchalski and Liu [22]). *For every $a \in (0,1]$ and $d \in \mathbb{N}$, there is $b = b(d,a) \in (0,1]$ with the following property. Let $\mathcal{C}$ be a collection of n convex sets in $\mathbb{R}^d$ $(n \geq d+1)$. Assume that the number of $(d+1)$-tuples with a nonempty intersection is at least $a\binom{n}{d+1}$. Then there is a point common to at least bn sets in $\mathcal{C}$.*

The largest possible value for $b(d,a)$ is $1-(1-a)^{1/(d+1)}$ due to Kalai [17] and Eckhoff [8] (i.e., it is even known that b cannot be larger). We remark that the Helly theorem is a special case when setting $a = 1$.

There is also a topological extension of the fractional Helly theorem by Alon, Kalai, Matoušek, and Meshulam [2] (with the same bound for b). They actually prove a bit stronger result [in order to obtain a topological (p,q)-theorem]; however, we prefer to avoid the technical details, and so we present the result only in this simpler form.

Theorem 6.8. *For every $a \in (0,1]$ and $d \in \mathbb{N}$, there is $b = b(d,a) \in (0,1]$ with the following property. Let K be a d-Leray complex with n vertices $(n \geq d+1)$. Assume that there are at least $a\binom{n}{d+1}$ d-faces in K. Then there is a face of size at least $bn-1$ in K.*

The largest possible value for $b(d,a)$ is again $1-(1-a)^{1/(d+1)}$.

6.4 The (p,q) Theorem

Let p, q be integers such that $p \geq q \geq d+1$. A family $\mathcal{F}$ of convex sets in $\mathbb{R}^d$ has the (p,q) *property* if among every p sets of $\mathcal{F}$, some q have a nonempty intersection. The *pinning number*, $\pi(\mathcal{F})$, of a family $\mathcal{F}$ is the smallest number of points in $\mathbb{R}^d$ that intersect all members of $\mathcal{F}$.

Theorem 6.9 (*(p,q)*-theorem, Alon and Kleitman [1]). *For every $p \geq q \geq d+1$, there is a number $C = C(p,q,d)$ such that $\pi(\mathcal{F}) \leq C$ for every family of convex sets in $\mathbb{R}^d$ with the (p,q) property.*

The (p,q) theorem was originally conjectured by Hadwiger and Debrunner.

In order not to introduce a new symbol, we let C denote the smallest constant for which the assertion of the (p,q) theorem is valid. The Helly theorem simply says that $C(d+1,d+1,d) = 1$. In general, there are, however, big gaps between

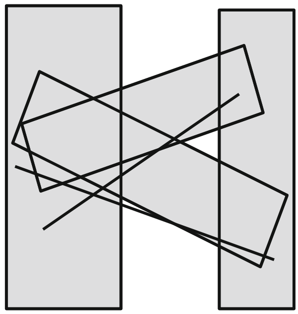


Fig. 6 Six convex sets with (4,3) property and pinning number 3

the lower and upper bounds for C. In the first nontrivial case, it is relatively easy to come up with an example showing $C(4,3,2) \geq 3$; see Fig. 6. Kleitman, Gyárfás, and Tóth [23] proved that $C(4,3,2) \leq 13$. The author of this survey believes that the actual value of $C(4,3,2)$ is much closer to 3 than 13; however, it seems difficult to obtain the precise value. The question of what the value of $C(p,q,d)$ is for larger p, q, and d is wide open.

Now we reformulate the setting for simplicial complexes. A simplicial complex K has the (p,q) *property* for $p \geq q$ if, among every p vertices of K, there are q of them forming a face (of dimension $q-1$). The *pinning number*, $\pi(K)$, of K is the smallest number of faces of K such that every vertex of K is in at least one of these faces. Then the statement of the (p,q) theorem remains valid even for d-Leray complexes and consequently for good covers due to Alon et al. [2].[3]

Theorem 6.10. *For every $p \geq q \geq d+1$, there is a number $C' = C'(p,q,d)$ such that $\pi(\mathsf{K}) \leq C'$ for every d-Leray complex K with the (p,q) property.*

6.5 *The Amenta Theorem*

In all previous cases, we were considering properties of collections of convex sets. Now we replace *convex sets* with *a finite disjoint union of convex sets*. It turns out that there is also a Helly-type theorem for this case (if we keep this property for intersections also).

Let $\mathcal{F}$ be a finite family of subsets of some ground set. A family $\mathcal{G}$ is an $(\mathcal{F},k)$-*family* if, for every nonempty $\mathcal{G}' \subseteq \mathcal{G}$, the intersection of elements of $\mathcal{G}'$ is the disjoint union of at most k members of $\mathcal{F}$. We have defined the Helly number only for simplicial complexes. For purposes of this subsection, we say that a family $\mathcal{F}$ has *Helly number* $h = h(\mathcal{F})$ equal to the Helly number of the nerve of $\mathcal{F}$. (Here we allow $\mathcal{F}$ to be possibly infinite.)

[3]Precisely speaking, Alon et al. state the theorem for good covers only; however, the same reasoning can be used for d-Leray complexes.

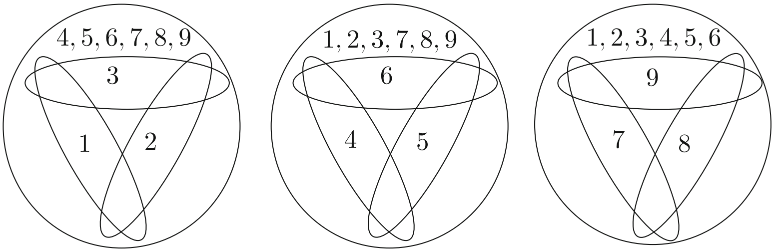


Fig. 7 The lower bound for Amenta's theorem. The convex hulls are a bit enlarged in order to make the figure more lucid. Moreover, the notation of the sets is simplified

Theorem 6.11 (Amenta [5]). *Let $\mathcal{F}$ be a finite family of compact convex sets in $\mathbb{R}^d$. Let $\mathcal{G}$ be an $(\mathcal{F},k)$-family. Then $h(\mathcal{G}) \leq k(d+1)$.*

The bound $k(d+1)$ in Amenta's theorem is optimal, as can be shown with the following example. Let $\mathcal{C} = \{C_1,\ldots,C_{d+1}\}$ be a collection of convex sets in $\mathbb{R}^d$ such that every d of them has a nonempty intersection; however, the intersection of the whole collection is empty (for example, $\mathcal{C}$ might be the collection of facets of a d-simplex). Let $\mathcal{C}_* = \mathcal{C} \cup \{C\}$, where C is the convex hull of the union of sets in $\mathcal{C}$. Let's consider k disjoint copies $\mathcal{C}_*^i = \{C_1^i,\ldots,C_{d+1}^i,C^i\}$ of $\mathcal{C}_*$ for $i \in \{1,\ldots,k\}$ such that the sets C^i are pairwise disjoint. Now we construct sets D_j^i for $i \in \{1,\ldots,k\}$, $j \in \{1,\ldots,d+1\}$ by setting

$$D_j^i = \left(\bigcup_{m\neq i} C^m\right) \cup C_j^i.$$

Then $\mathcal{G} = \{D_j^i\}$ is an $(\mathcal{F},k)$-family and $h(\mathcal{G}) = k(d+1)$. See Fig. 7.

Now we focus on a topological version. Due to the fact that we consider the union of sets, there is no simple statement for a topological generalization of Amenta's theorem using d-Leray complexes. Thus, we prefer to set up the statement for good covers in this case.

Theorem 6.12 (Kalai, Meshulam [21]). *Let $\mathcal{F}$ be a finite good cover in $\mathbb{R}^d$. Let $\mathcal{G}$ be an $(\mathcal{F},k)$-family. Then $h(\mathcal{G}) \leq k(d+1)$.*

Eckhoff and Nischke [10] recently proved Amenta's theorem in a very abstract setting via Morris's pigeonhole principle. A (possibly infinite) family $\mathcal{F}$ is *intersectional* if, for any finite subfamily $\mathcal{F}'$, the intersection $\cap\mathcal{F}'$ either is empty or belongs to $\mathcal{F}$. We call $\mathcal{F}$ *nonadditive* if, for any finite subfamily $\mathcal{F}'$ of disjoint sets (at least two nonempty), $\cup\mathcal{F}' \notin \mathcal{F}$. The following theorem generalizes the previous two theorems by setting $\mathcal{F}$ to be either a family of convex compact sets, or a family of all intersections of a good cover.

Theorem 6.13. *Let $\mathcal{F}$ be an intersectional and nonadditive set family. If $\mathcal{G}$ is an $(\mathcal{F},k)$ family, then $h(\mathcal{G}) \leq kh(\mathcal{F})$.*

From another point of view, a generalization of Theorem 6.12 for collections of sets in more general topological spaces (than $\mathbb{R}^d$) was obtained by Colin de Verdière, Ginot, and Goaoc [7]. More importantly, their generalization applies to collections of sets that need not come from good covers. For instance, it applies to a collection of two sets homeomorphic to a ball which intersect in two balls.

By Γ, we denote a locally arc-wise connected topological space. Then d_Γ is the smallest integer such that every open subset of Γ has a trivial $\mathbb{Q}$-homology in dimension d_Γ and higher. An open subset of Γ with singular $\mathbb{Q}$-homology equivalent to a point is a $\mathbb{Q}$-*homology cell*. A family of open subsets of Γ is acyclic if, for every nonempty subfamily $\mathcal{G} \subset \mathcal{F}$, the intersection of $\mathcal{G}$ is a disjoint union of $\mathbb{Q}$-homology cells. Colin de Verdière et al. proved the following result.

Theorem 6.14. *Let $\mathcal{F}$ be a finite acyclic family of open subsets of locally arc-wise connected topological space Γ. If any subfamily of $\mathcal{F}$ intersects in at most k connected components, then the Helly number of $\mathcal{F}$ is at most $k(d_\Gamma + 1)$.*

In particular, $d_{\mathbb{R}^d} = d$ and every good cover is acyclic; therefore, Theorem 6.14 indeed generalizes Theorem 6.12. Let's also remark that $d_\Gamma = d$ for a d-dimensional manifold that is either noncompact or nonorientable, and $d_\Gamma = d + 1$ for a compact orientable manifold.

Theorem 6.14 can be further generalized when the homology is zero only from a certain dimension. For this case, we already refer the reader to [7]. On the other hand, the case when the homology vanishes for small dimensions has already been considered in [28]. The bound to the Helly number is, however, much weaker.

Acknowledgements The author thanks Xavier Goaoc for discussions on new generalizations of Amenta's theorem and additional remarks, Jiří Matoušek for many valuable comments to the preliminary version of the survey, and also Janos Pach for discussions about d-dimensional complexes that are not $2d$-representable.

References

1. N. Alon, D. Kleitman, Piercing convex sets and the Hadwiger Debrunner (p,q)-problem. Adv. Math. **96**(1), 103–112 (1992)
2. N. Alon, G. Kalai, J. Matoušek, R. Meshulam, Transversal numbers for hypergraphs arising in geometry. Adv. Appl. Math. **130**, 2509–2514 (2002)
3. N. Alon, Expanders, sorting in rounds and superconcentrators of limited depth, in *Proceedings of the 17th ACM Symposium on Theory of Computation*, 1985, pp. 98–102 http://dl.acm.org/citation.cfm?id=22156
4. N. Alon, Eigenvalues, geometric expanders, sorting in rounds, and Ramsey theory. Combinatorica **6**(3), 207–219 (1986)
5. N. Amenta, A short proof of an interesting Helly-type theorem. Discrete Comput. Geom. **15**, 423–427 (1996)
6. L.J. Billera, A. Björner, Face numbers of polytopes and complexes, in *Handbook of Discrete and Computational Geometry*. CRC Press Ser. Discrete Math. Appl. (CRC Press, Boca Raton, FL, 1997), pp. 291–310

7. É. Colin de Verdière, G. Ginot, X. Goaoc, Helly numbers of acyclic families. Preprint; http://arxiv.org/abs/1101.6006, 2011
8. J. Eckhoff, An upper-bound theorem for families of convex sets. Geom. Dedicata **19**(2), 217–227 (1985)
9. J. Eckhoff, in *Helly, Radon and Carathéodory type theorems*, ed. by P.M. Gruber, J.M. Wills. Handbook of Convex Geometry (North-Holland, Amsterdam, 1993)
10. J. Eckhoff, K.-P. Nischke, Morris's pigeonhole principle and the Helly theorem for unions of convex sets. Bull. Lond. Math. Soc. **41**(4), 577–588 (2009)
11. A. Flores, Über n-dimensionale Komplexe die im R_{2n+1} absolut selbstverschlungen sind. Ergeb. Math. Kolloq. **4**, 6–7 (1932/1934)
12. J.E. Goodman, R. Pollack, R. Wenger, Geometric transversal theory, in *New Trends in Discrete and Computational Geometry*. Algorithms Combination, vol. 10 (Springer-Verlag, Berlin, 1993), pp. 163–198
13. J.E. Goodman, R. Pollack, R. Wenger, T. Zamfirescu, Arrangements and topological planes. Am. Math. Mon. **101**(9), 866–878 (1994)
14. A. Hatcher, *Algebraic Topology* (Cambridge University Press, Cambridge, 2001)
15. E. Helly, Über mengen konvexer Körper mit gemeinschaftlichen Punkten. Jahresber. Deustch. Math. Verein. **32**, 175–176 (1923)
16. E. Helly, Über Systeme von abgeschlossenen Mengen mit gemeinschaftlichen Punkten. Monaths. Math. Physik **37**, 281–302 (1930)
17. G. Kalai, Intersection patterns of convex sets. Isr. J. Math. **48**(2–3), 161–174 (1984)
18. G. Kalai, Algebraic shifting, in *Computational Commutative Algebra and Combinatorics (Osaka, 1999)*. Adv. Stud. Pure Math., vol. 33 (Math. Soc. Japan, Tokyo, 2002), pp. 121–163
19. G. Kalai, R. Meshulam, A topological colorful Helly theorem. Adv. Math. **191**(2), 305–311 (2005)
20. G. Kalai, R. Meshulam, Intersections of Leray complexes and regularity of monomial ideals. J. Combin. Theor. Ser. A **113**(7), 1586–1592 (2006)
21. G. Kalai, R. Meshulam, Leray numbers of projections and a topological Helly type theorem. J. Topol. **1**(3), 551–556 (2008)
22. M. Katchalski, A. Liu, A problem of geometry in R^n. Proc. Am. Math. Soc. **75**, 284–288 (1979)
23. D.J. Kleitman, A. Gyárfás, G. Tóth, Convex sets in the plane with three of every four meeting. Combinatorica, **21**(2), 221–232 (2001); Paul Erdős and His Mathematics (Budapest, 1999)
24. J. Kratochvíl, J. Matoušek, NP-hardness results for intersection graphs. Coment. Math. Univ. Carolin. **30**, 761–773 (1989)
25. J. Kratochvíl, J. Matoušek, Intersection graphs of segments. J. Comb. Theor. Ser. B **62**(2), 289–315 (1994)
26. C.G. Lekkerkerker, J. Ch. Boland, Representation of a finite graph by a set of intervals on the real line. Fund. Math. **51**, 45–64 (1962/1963)
27. L. Lovász, Problem 206. Matematikai Lapok **25**, 181 (1974)
28. J. Matoušek, A Helly-type theorem for unions of convex sets. Discrete Comput. Geom. **18**(1), 1–12 (1997)
29. J. Matoušek, *Lectures on Discrete Geometry* (Springer-Verlag, New York, 2002)
30. J. Matoušek, *Using the Borsuk-Ulam Theorem* (Springer, Berlin, 2003)
31. J. Matoušek, M. Tancer, Dimension gaps between representability and collapsibility. Discrete Comput. Geom. **42**(4), 631–639 (2009)
32. J.R. Munkres, *Elements of Algebraic Topology* (Addison-Wesley, Reading, MA, 1984)
33. J. Pach, A Tverberg-type result on multicolored simplices. Comput. Geom. **10**, 71–76 (1998)
34. G.Y. Perel'man, Realization of abstract k-skeletons of intersections of convex polyhedra in R^{2k-1}. *Geometric Questions in the Theory of Functions and Sets* (Kalinin. Gos. Univ., Kalinin, 1985), pp. 129–131 [in Russian]
35. M. Tancer, d-collapsibility is NP-complete for d greater or equal to 4. Chicago J. Theoret. Comput. Sci. **2010**(3), 1–28 (2010)
36. M. Tancer, Non-representability of finite projective planes by convex sets. Proc. Am. Math. Soc. **138**(9), 3285–3291 (2010)

37. M. Tancer, A counterexample to Wegner's conjecture on good covers. *Discrete & Computational Geometry* (2012), **47**(2), pp. 266–274. http://www.springerlink.com/content/nn36q661u87n111u/.
38. M. Tancer, *d*-representability of simplicial complexes of fixed dimension. in *J. Comput. Geom.* **2**(1), **1** pp. 183–188 (2011) http://arxiv.org/abs/1107.1170
39. R.E. van Kampen, Komplexe in euklidischen Räumen. Abh. Math. Sem. Hamburg **9**, 72–78 (1932) Berichtigung dazu, *ibid*. (1932) 152–153
40. G. Wegner, Eigenschaften der Nerven homologisch-einfacher Familien im R^n, PhD thesis, Universität Göttingen, 1967, [in German]
41. G. Wegner, *d*-collapsing and nerves of families of convex sets. Arch. Math. **26**, 317–321 (1975)

Construction of Locally Plane Graphs with Many Edges

Gábor Tardos

Abstract A graph drawn in the plane with straight-line edges is called a geometric graph. If no path of length at most k in a geometric graph G is self-intersecting, we call G k-locally plane. The main result of this chapter is a construction of k-locally plane graphs with a superlinear number of edges. For the proof, we develop randomized thinning procedures for edge-colored bipartite (abstract) graphs that can be applied to other problems as well.

1 Introduction

A *geometric graph* G is a straight-line drawing of a simple, finite (abstract) graph (V,E); i.e., we identify the vertices $x \in V$ with distinct points in the Euclidean plane, and we identify any edge $\{x,y\} \in E$ with the straight-line segments xy in the plane. We assume that the edge xy does not pass through any vertex of G besides x and y. We call (V,E) the abstract graph *underlying* G. We say that the edges $e_1, e_2 \in E$ *cross* if the corresponding line segments cross each other, i.e., if they have a common interior point. We say that a subgraph of G is *self-intersecting* if it contains a pair of crossing edges.

Geometric graphs without crossing edges are plane drawings of planar graphs: They have at most $3n-6$ edges if $n \geq 3$ is the number of vertices.

Avital and Hanani [3], Erdős, and Perles initiated in the mid-1960s the systematic study of similar questions for more complicated *forbidden configurations*: Let H be a set of forbidden configurations (geometric subgraphs). What is the maximal number of edges of an n vertex geometric graph not containing any configuration belonging to H? This problem can be regarded as a geometric version of the

G. Tardos (✉)
Rényi Institute, Budapest, Hungary; and School of Computing Science,
Simon Fraser University, Burnaby, BC, Canada
e-mail: tardos@renyi.hu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_29,

fundamental problem of extremal graph theory: What is the maximum number of edges that an abstract graph on n vertices can have without containing subgraphs of a certain kind?

Many questions of the above type on geometric graphs have been addressed in recent years. In a number of papers, linear upper bounds have been established for the number of edges of a graph, under various forbidden configurations. They include the configurations of three pairwise crossing edges [2], four pairwise crossing edges [1], the configurations of an edge crossed by many edges [9], or even two large stars with all edges of one of them crossing all edges of the other [13].

For a constant number of 5 or more pairwise crossing edges, Pavel Valtr has the best result [11]: A geometric graph on n vertices avoiding this configuration has $O(n \log n)$ edges, but no construction is known with a superlinear number of edges. Adam Marcus and the present author [4] building on an earlier result of Pinchasi and Radoičić [10] prove an $O(n^{3/2} \log n)$ bound on the number of edges of an n vertex geometric graph not containing self-intersecting cycles of length 4. No construction is known better than the $O(n^{3/2})$ edges an abstract graph having no cycles of length 4 can have.

For surveys on geometric graph theory, consult [5, 6].

In this chapter, we consider forbidding self-intersecting paths. For $k \geq 3$, we call a geometric graph *k-locally plane* if it has no self-intersecting subgraph (whose underlying abstract graph is) isomorphic to a path of length at most k.

Pach et al. [7] consider 3-locally plane graphs, i.e., the case of geometric graphs with no self-intersecting paths of length 3. They prove matching lower and upper bounds of $\Theta(n \log n)$ on the maximal number of edges of a 3-locally plane graph on n vertices.

We extend the lower-bound result of [7] by considering self-intersecting drawings of longer paths as forbidden configurations. Technically, k-locally plane graphs are defined by forbidding self-intersecting paths of length *k or shorter*, but forbidding only self-intersecting paths of length *exactly* k would lead to almost the same extremal function. Indeed, one can delete at most nk edges from any graph on n vertices, such that all the nonzero degrees in the remaining graph are larger than k. This ensures that all shorter paths can be extended to a path of length k. It is possible, but not likely, that if one only forbids paths of length k with *the first and last edges crossing*, a significantly higher number of edges is achievable.

For even k, a geometric graph is k-locally plane if and only if the $k/2$-neighborhood of any vertex x is intersection-free. Note that this requirement is much stronger than the similar condition on abstract graphs, namely, that the $k/2$ neighborhood of any point is planar. One can construct graphs with girth larger than k and $\Omega(n^{\frac{k}{k-1}})$ edges. In such a graph, the $k/2$-neighborhood of any vertex is a tree, but by [7], the graph does not even have 3-locally plane drawing.

Extending the lower bound result in [7], we prove in Theorem 3 that for arbitrary fixed $k \geq 3$, there exist k-locally plane graphs on n vertices with $\Omega(n \log^{(\lfloor k/2 \rfloor)} n)$

edges. Here $\log^{(t)}$ denotes a t-times iterated logarithm, and the hidden constant in Ω depends on k. Given two arbitrarily small disks in the plane, we can even ensure that all edges of the constructed graph connect a vertex from the first disk with another vertex from the second. This ensures that all the edges of the constructed geometric graph are arbitrarily close to each other in length and direction. In the view of the author, this makes the existence of a high average degree (or, for that matter, a high minimum degree) 100-locally plane graphs even more surprising.

As a simple corollary, we can characterize the abstract graphs H such that any geometric graph having no self-intersecting subgraph isomorphic to H has a linear number of edges. These graphs H are the forests with at least two nontrivial components. To see the linear bound for the number of edges of a geometric graph avoiding a self-intersecting copy of such a forest H, first delete a linear number of edges from an arbitrary geometric graph G until all nonzero degrees of the remaining geometric graph G' are at least $|V(H)|$. If G' is crossing-free, the linear bound of the number of edges follows. If you find a pair of crossing edges in G', they can be extended to a subgraph of G' isomorphic to H. On the other hand, if H contains a cycle, then even an abstract graph avoiding it can have a superlinear number of edges. If H is a tree of diameter k, then a k-locally plane geometric graph has no self-intersecting copy of H. Notice that the extremal number of edges in this case (assuming $k > 2$) is $O(n \log n)$ by [7], thus much smaller than the $\Omega(n^{\alpha})$ edges ($\alpha > 1$) for forbidden cycles.

The main tool used in the proof of the above result is a randomized thinning procedure that takes a d-edge-colored bipartite graph of average degree $\Theta(d)$ and returns a subgraph on the same vertex set with average degree $\Theta(\log d)$ that does not have a special type of colored path (walk) of length 4. The procedure can be applied recursively to obtain a subgraph avoiding longer paths of certain types. We believe this thinning procedure to be of independent interest. In particular, it can be used to obtain optimal 0–1 matrix constructions for certain avoided submatrix problems; see the exact statement in Sect. 4 and the details in [12].

In Sect. 2, we define two thinning procedures for edge-colored bipartite graphs and prove their main properties. This is the most technical part of the chapter. While these procedures proved useful in other settings too (and the author finds the involved combinatorics appealing), this entire section can be skipped if one reads the definition of k-flat graphs (the two paragraphs before Lemma 2.14) and is willing to accept Corollary 2.1 at the end of the section (we also use the simple observation in Lemma 2.15). In fact, in order to understand the main ideas behind the main result of this chapter, it is recommended to skip Sect. 2 on the first reading and to go straight to Sect. 3, where we use Corollary 2.1 to construct locally plane graphs with many edges. In Sect. 4, we comment on the optimality of the thinning procedures and have some concluding remarks.

2 Thinning

In this section, we state and prove combinatorial statements about edge-colored abstract graphs; i.e., we do not consider geometric graphs here at all. The connection to locally plane geometric graphs will be made clear in Sect. 3.

A *bipartite graph* is a triple $G=(A,B,E)$ with disjoint vertex sets A and B (called *sides*) and edge set $E\subseteq A\times B$. In particular, all graphs considered in this chapter are *simple*; i.e., they do not have multiple edges or loops. The edge connecting the vertices x and y of G is denoted by xy or (x,y). The latter notation is only used if $x\in A$ and $y\in B$. By a *d-edge coloring* of a graph, we mean a mapping $\chi:E\to\{1,2,\ldots,d\}$ such that adjacent edges receive different colors. When we do not specify d, we call such coloring a *proper edge coloring*, but we always assume that the "set of colors" is linearly ordered. The degree of any vertex in G is at most the number d of colors, and our results are interesting if the average degree is close to d. Unless stated otherwise, the subgraphs of an edge-colored graph are considered with the inherited edge coloring. Our goal is to obtain a subgraph of G with as many edges as possible without containing a certain type of colored path or walk.

2.1 Heavy Paths

A simple example of the above concept is the following. We call a path $P=v_0v_1v_2v_3$ of length 3 *heavy* if $v_0\in B$ and the colors $c_1=\chi(v_0v_1)$, $c_2=\chi(v_1v_2)$, $c_3=\chi(v_2v_3)$ satisfy $c_2<c_1\le c_3$. The next lemma describes a thinning procedure that gets rid of heavy paths. Although we do not need this lemma in our construction, we present it as a simple analogue of our results for more complicated forbidden walks.

Lemma 2.6. *Let $G=(A,B,E)$ be a bipartite graph with a proper edge coloring $\chi:E\to\{1,2,\ldots,d\}$. Then there exists a subgraph $G'=(A,B,E')$ of G with $|E'|\ge|E|/(3\lceil\sqrt{d}\,\rceil)$ that does not contain heavy paths.*

The constant 3 in the lemma could be replaced by the base of the natural logarithm. Notice that if G had average degree $\Theta(d)$, then the average degree of G' is $\Omega(\sqrt{d})$.

Proof. Let $t=\lceil\sqrt{d}\,\rceil$ and select a uniform random value $i_y\in\{1,2,\ldots,t\}$ independently for each vertex $y\in B$. We say that an edge $e\in E$ is of *class* $\lceil\chi(e)/t\rceil$. We call an edge $e=(x,y)\in E$ *eligible* if its class is i_y. Let the subgraph $G'=(A,B,E')$ consist of those eligible edges $e=(x,y)\in E$ for which there exists no other eligible edge $(x,y')\in E$ of the same class. Note that the words "class" and "eligible" will be used in a different meaning when defining the two thinning procedures in the next subsection.

By the construction, all edges incident to a vertex $x\in A$ have different classes, and all edges incident to a vertex $y\in B$ have the same class. Let e_1, e_2, and e_3 form a path in G' starting in B. Then e_2 and e_3 are of the same class, while the class of e_1 is different. For the colors $c_i = \chi(e_i)$, this rules out the order $c_2 < c_1 \leq c_3$. Thus, G' does not contain a heavy path. Note that another order, $c_3 \leq c_1 < c_2$, is also impossible.

The number of edges in G' depends on the random choices we made. Any edge $(x,y) \in E$ is eligible with probability $1/t$, and this is independent for all the edges incident to a vertex $x \in A$. As edges of a fixed color form a matching, there are at most t edges of any given class incident to x. Thus, we have

$$\Pr[(x,y) \in E'] \geq \frac{(1-1/t)^{t-1}}{t} > \frac{1}{3t}.$$

The expected number of edges in G' is

$$\mathrm{Exp}[|E'|] \geq \frac{|E|}{3t}.$$

It is possible to choose the random variables i_y so that the size of E' is at least as large as its expected value. This proves the lemma. □

2.2 *Fast and Slow Walks*

Next, we turn to more complicated forbidden subgraphs. For motivation, we mention that self-crossing paths of length 4 in the 3-locally plane graphs of [7] (considered with their natural edge coloring) are exactly the *fast walks* (to be defined below). For technical reasons, it will be more convenient to consider walks, i.e., to permit that a vertex is visited more than once, but we will not allow *backtracking*, i.e., turning back on the same edge immediately after it was traversed. Thus, for us, a *walk* of length k is a sequence $v_0, v_1, \ldots, v_k$ of vertices in the graph such that $v_{i-1}v_i$ is an edge for $1 \leq i \leq k$ and $v_{i-2} \neq v_i$ for $2 \leq i \leq k$. The χ*-coloring* (or simply coloring) of this walk is the sequence $(\chi(v_0v_1), \chi(v_1v_2), \ldots, \chi(v_{k-1}v_k))$ of the colors of the edges of the walk. If χ is a proper edge coloring, then any two consecutive elements of the coloring sequence are different.

We use log to denote the binary logarithm. We introduce the notation $P(a,b)$ for two nonequal strings $a,b \in \{0,1\}^t$ to denote the first position $i \in \{1,2,\ldots,t\}$, where a and b differ. We consider the set $\{0,1\}^t$ to be ordered lexicographically; i.e., for $a,b \in \{0,1\}^t$, we have $a < b$ if a has 0 in position $P(a,b)$ (and thus b has 1 there).

The following trivial observation is used often in this chapter. We state it here without a proof.

Lemma 2.7. *Let $t \geq 1$ and let a, b, and c be distinct binary strings of length t with $P(a,b) < P(a,c)$. We have $P(b,c) = P(a,b)$. Furthermore, $a > b$ implies $c > b$, and $a < b$ implies $c < b$.*

A walk of length 4 with coloring (c_1,c_2,c_3,c_4) is called a *fast walk* if $c_2 < c_3 < c_4 \leq c_1$. Note that a fast walk may start in either class A or B. We call a walk of length 4 a *slow walk* if it starts in the class B and its coloring (c_1,c_2,c_3,c_4) satisfies $c_2 < c_3 < c_4$ and $c_2 < c_1 \leq c_4$. Note that either the color c_1 or c_3 can be larger in a slow walk, or they can be equal.

The two *thinning procedures* below find a random subgraph of an edge-colored bipartite graph. One is designed to avoid slow walks, while the other is designed to avoid fast walks.

Lexicographic thinning. Let $G = (A,B,E)$ be a bipartite graph and let $\chi : E \to \{1,\dots,d\}$ be a proper edge coloring with $d \geq 2$. *Lexicographic thinning* is a randomized procedure that produces a subset $E' \subseteq E$ of the edges and the corresponding subgraph $G' = (A,B,E')$ of G as follows:

Let $t = \left\lceil \frac{\log d}{2} \right\rceil + 1$. Let H be the set of triplets (a,i,z), where $a \in \{0,1\}^t$, $i \in \{2,3,\dots,t\}$, $z \in \{1,2,3,\dots,2^i\}$, and a has 0 in position i. Straightforward calculation gives that $|H| = 2^{2t} - 2^{t+1} \geq 2d$.

We order H lexicographically; i.e., $(a,i,z) < (b,j,s)$ if $a < b$, or $a = b$ and $i < j$, or $(a,i) = (b,j)$ and $z < s$.

Consider the following random function $F : \{1,\dots,d\} \to H$. We select uniformly at random the value $F(1) = (a,i,z) \in H$ with the property that the first bit of a is 0. We make $F(2)$ to be the next element of H larger than $F(1)$, and in general $F(k)$ is the next element of H larger than $F(k-1)$ for $2 \leq k \leq d$. As $|H| \geq 2d$ and $F(1)$ is chosen from the first half of H, this defines F. In what follows, we simply identify the color k with the element $F(k) \in H$ without any reference to the function F.

We say that $(a,i,z) \in H$ and any edge with this color is of *class* a and *type* i, while z will play no role except in counting how many values it can take.

We choose an independent uniform random value $a_x \in \{0,1\}^t$ for each vertex $x \in A \cup B$. Let $e = (x,y) \in E$ be an edge of class a and type i. We say that e is *eligible* if $a = a_y < a_x$ and $P(a,a_x) = i$. Let the subgraph $G' = (A,B,E')$ contain those edges $e \in E$ that are eligible but not adjacent to another eligible edge e' of the same type as e.

Reversed thinning. Let $G = (A,B,E)$ be a bipartite graph and let $\chi : E \to \{1,\dots,d\}$ be a proper edge coloring with $d \geq 2$. *Reversed thinning* is a randomized procedure that produces a subset $E' \subseteq E$ of the edges and the corresponding subgraph $G' = (A,B,E')$ of G.

Reversed thinning is almost identical to lexicographic thinning; the only difference is in the ordering of the set H. We define t and H as in the case of lexicographic thinning. Recall that H is the set of triplets (a,i,z), where $a \in \{0,1\}^t$, $i \in \{2,3,\dots,t\}$, $z \in \{1,2,\dots 2^i\}$, and a has 0 in position i. We still order $\{0,1\}^t$ lexicographically, but now we reverse the lexicographic order of H in the middle term i. That is, we have $(a,i,z) < (b,j,s)$ if $a < b$, or $a = b$ and $i > j$ or $(a,i) = (b,j)$ and $z < s$.

We define the function $F : \{1,\ldots,d\} \to H$, the *types* and *classes* of colors and edges, *eligible* edges, and the subset E' of edges the same way as for lexicographic thinning, but using this modified ordering of H.

Note that we associated a type in $\{2,\ldots,t\}$ and a class in $\{0,1\}^t$ to each edge in either procedure and they satisfy the following:

- An edge with smaller class has smaller color.
- Among edges of equal class, an edge with smaller type has smaller color in the case of lexicographic thinning, and it has larger color in the case of reversed thinning.
- Among the edges incident to a vertex, at most 2^i have the same class a and the same type i.

Proving most of the properties of the thinning procedures, this is all we need to know about how classes and types are associated with the edges, and we could use a deterministic scheme for F. But for Lemma 2.11, we need that all types are well represented; the randomization in the identification function F (as well as the dummy first bit of the class) is introduced to ensure this on the average. This randomization is not needed if one assumes all color classes have roughly the same size.

The next lemmas state the basic properties of the thinning procedure. Lemma 2.8 lists common properties of the two procedures, while Lemmas 2.9 and 2.10 state the result of the thinning satisfies its "design criteria," avoiding slow or fast walks. Finally, Lemma 2.11 shows that enough edges remain in the constructed subgraphs on average. Note that Lemmas 2.9 and 2.10 are special cases of the more complex Lemmas 2.12 and 2.13 proved independently. We state and prove the simple cases separately for clarity, but these proofs could be skipped.

Lemma 2.8. *Let $G = (A,B,E)$ be a bipartite graph with a proper edge coloring $\chi : E \to \{1,\ldots,d\}$. If $G' = (A,B,E')$ is the result of either the lexicographic or the reversed thinning, then we have*

a) Adjacent edges in G' have distinct types.
b) If two edges of G' meet in B, they have the same class.
c) Suppose two distinct edges e and e' of G' meet in A. Let their classes and types be a, a' and i, i', respectively. If $i < i'$, then $a < a'$ and $P(a,a') = i$.
d) G' has no heavy path.

Proof. The definition of E' immediately gives (a).

For (b), note that all eligible edges incident to $y \in B$ have a_y for class.

For (c), let $x \in A$ be the common vertex of the two edges and apply Lemma 2.7 for a_x, a, and a'.

Finally, (d) follows since if a walk of G' starts in B, then its coloring (c_1,c_2,c_3) must satisfy that c_1 and c_2 have different class by (c), but c_2 and c_3 have the same class by (b), so $c_2 < c_1 \leq c_3$ is impossible. □

Lemma 2.9. *Let $G = (A,B,E)$ be a bipartite graph with a proper edge coloring. Lexicographic thinning produces a subgraph G' with no slow walk.*

Proof. Suppose $v_0v_1v_2v_3v_4$ is a walk in G' starting at $v_0 \in B$ and let its coloring be (c_1,c_2,c_3,c_4). Assume $c_1 > c_2 < c_3 < c_4$. We show that $c_1 > c_4$, so this walk is not slow. By Lemma 2.8(a, b), as c_2 and c_3 are colors of edges incident to $v_2 \in B$, they have the same class, but they have different types: $c_2 = (a,i,z)$, $c_3 = (a,j,s)$ with $i \neq j$. We use lexicographic ordering, so $c_2 < c_3$ implies $i < j$. Both c_1 and c_2 are colors of edges incident to $v_1 \in A$, so by Lemma 2.8(c), their classes are different. Since $c_1 > c_2$, we have $b > a$ for the class b of c_1. Still by Lemma 2.8(c), $P(a,b) = i$. Similarly, c_3 and c_4 are colors of distinct edges in E' incident to $v_3 \in A$, so they have different classes. As $c_3 < c_4$, we have $c > a$ for the class c of c_4. We have $P(a,c) = j$. By Lemma 2.7, we have $b > c$. This proves $c_1 > c_4$, as claimed. □

Lemma 2.10. *Let $G = (A,B,E)$ be a bipartite graph with a proper edge coloring. Reversed thinning produces a subgraph G' with no fast walk.*

Proof. Suppose $v_0v_1v_2v_3v_4$ is a walk in G' with coloring (c_1,c_2,c_3,c_4). Assume $c_1 > c_2 < c_3 < c_4$. We show that $c_1 < c_4$, so this walk is not fast. First assume the walk starts at $v_0 \in A$. As G' does not contain a heavy path, $v_3v_2v_1v_0$ is not heavy, so $c_1 < c_3$. This implies $c_1 < c_4$, as claimed.

Now assume $v_0 \in B$. Just as in the proof of the previous lemma, c_2 and c_3 are colors of edges incident to $v_2 \in B$, so they have the same class, but they have different types: $c_2 = (a,i,z)$, $c_3 = (a,j,s)$ with $i \neq j$. We use the reversed ordering, so $c_2 < c_3$ implies $i > j$. Both c_1 and c_2 are colors of edges incident to $v_1 \in A$, so their classes are different. Since $c_1 > c_2$, we have $b > a$ for the class b of c_1 and $P(a,b) = i$. Similarly, c_3 and c_4 are colors of distinct edges in E' incident to $v_3 \in A$, so they have different classes. As $c_3 < c_4$, we have $c > a$ for the class c of c_4 and $P(a,c) = j$. By Lemma 2.7, we have $c > b$. This proves $c_1 < c_4$, as claimed. □

Below we estimate the number of edges in E'. Recall that both thinning procedures are randomized. We can show that the subgraphs they produce have a large expected number of edges. We did not make any effort to optimize for the constant in this lemma.

Lemma 2.11. *Let $G = (A,B,E)$ be a bipartite graph with a d-edge coloring. Let $G' = (A,B,E')$ be the result of either the lexicographic or the reversed thinning. We have*

$$\mathrm{Exp}[|E'|] \geq \frac{t-1}{240d}|E| \geq \frac{\log d}{480d}|E|.$$

Proof. We compute the probability for a fixed edge $e = (x,y) \in E$ to end up in E'. For this we break down the random process producing E' into three phases. In the first phase, we select F. With F, the color $\chi(e)$ is identified with an element of H; most importantly, the type of e is fixed. In the second phase, we select a_x and a_y uniformly at random. These choices determine if e is eligible. If e is not eligible, then $e \notin E'$. So in the third phase, we consider F, a_x, and a_y fixed and assume e is eligible. We select the random values a_z for vertices $z \neq x,y$. This affects if other edges are eligible and if $e \in E'$. Here is the detailed calculation:

Let $e \in E$ have the color $\chi(e) = k \in \{1,2,\ldots,d\}$. The choice of F in the first phase determines $F(k) = (a,i,z) \in H$. By the construction of F, if we call a' the last $t-1$ bits of a, then (a',i,z) is uniformly distributed among all its possible values. In particular, the probability that e becomes a type-i edge is exactly

$$\Pr[e \text{ is of type } i] = \frac{2^{t+i-2}}{2^{2t-1}-2^t} = \frac{2^i}{2^{t+1}-4}.$$

For the second phase, we consider the function F identifying colors with elements of H fixed. Consider an edge $e = (x,y)$ of color $(a,i,z) \in H$. This edge is eligible if $a_y = a < a_x$ and $P(a,a_x) = i$. This determines a_y and the first i bits of a_x. Recall that by the definition of H, the string a has 0 in position i. Thus, the probability that the edge e of type i is eligible is exactly 2^{-t-i}.

Assume for the third phase that e is eligible. Consider another edge $e' = (x,y') \in E$ with color $\chi(e') = (a',i,z')$ of type i. If a and a' do not agree in the first i positions, then e' is not eligible. If they agree in the first i positions, then e' is eligible if and only if $a_{y'} = a'$, so with probability 2^{-t}. Let k'_e be the number of edges (x,y') of type i with the first i digits of their class agreeing with a, but not counting e itself. We have $k'_e < 2^t$.

Consider now an edge $e'' = (x'',y) \in E$ with color (a'',i,z'') such that $e'' \neq e$. As e is eligible, e'' can only be eligible if $a'' = a$. If $a'' = a$, then e'' is eligible if and only if a_x and $a_{x''}$ agree in the first i digits. This happens with probability 2^{-i}. For the number k''_e of the edges $(x'',y) \neq e$ of type i and class a, we have $k''_e < 2^i$.

In phase three, the eligibilities of all the edges e' and e'' adjacent to e are independent events.

Still consider the function F fixed. The total probability for an edge e of type i to be in E' is

$$\begin{aligned}\Pr[e \in E' | F] &= 2^{-t-i}(1-2^{-t})^{k'_e}(1-2^{-i})^{k''_e} \\ &\geq 2^{-t-i}(1-2^{-t})^{2^t-1}(1-2^{-i})^{2^i-1} \\ &> 2^{-t-i}/7.5.\end{aligned}$$

The total probability of $e \in E'$ can be calculated from the distribution of its type and the above conditional probability depending on its type:

$$\Pr[e \in E'] > \sum_{i=2}^{t} \frac{2^i}{2^{t+1}-4} \cdot \frac{2^{-t-i}}{7.5} > \frac{t-1}{15 \cdot 2^{2t}} > \frac{t-1}{240d}.$$

The expected number of edges in E' is then

$$\operatorname{Exp}[|E'|] > \frac{t-1}{240d}|E| \geq \frac{\log d}{480d}|E|. \qquad \square$$

Theorem 1. *Let $G = (A,B,E)$ be a bipartite graph with a proper edge coloring. There exists a subgraph $G' = (A,B,E')$ of G without a slow walk and with $|E'| > \frac{\log d}{480d} \cdot |E|$. Similarly, there exists a subgraph $G'' = (A,B,E'')$ of G without a fast walk and with $|E''| > \frac{\log d}{480d} \cdot |E|$.*

Proof. By Lemmas 2.9 and 2.10, the results of the lexicographic and reversed thinnings avoid the slow and fast walks, respectively. There exists an instance of the random choices with the size of E' being at least its expectation given in Lemma 2.11. This proves the theorem. □

2.3 Longer Forbidden Walks

Here we generalize the concept of fast and slow walks to longer walks. Consider a bipartite graph $G = (A, B, E)$ with a proper edge coloring. For $k \geq 2$, we call a walk of length $2k$ in G a *k-fast walk* if its coloring $(c_1, \dots, c_{2k})$ satisfies $c_1 > c_2 > \cdots > c_k < c_{k+1} < c_{k+2} < \cdots < c_{2k}$ and $c_1 \geq c_{2k}$. For $k \geq 2$, we call a walk of length $2k$ in G a *k-slow walk* if its coloring $(c_1, \dots, c_{2k})$ satisfies the following: $c_{2j-1} > c_{2j}$ for $1 \leq j \leq k/2$; $c_{2j} < c_{2j+1}$ for $1 \leq j < k/2$; $c_{2j-1} < c_{2j}$ for $k/2 < j \leq k$; $c_{2j} > c_{2j+1}$ for $k/2 \leq j < k$; and finally, $c_1 \geq c_{2k}$. If a k-slow walk starts in the vertex set B, we call it a (k, B)-slow walk; otherwise, it is a (k, A)-slow walk.

Notice that 2-fast walks are the fast walks, and $(2, B)$-slow walks are the slow walks with their orientation reversed. For the coloring $(c_1, \dots, c_{2k})$ of a k-fast walk, c_j is in between c_{j-1} and c_{j+1} for all $1 < j < 2k$, $j \neq k$, while c_k is the smallest color on this list. For the coloring $(c_1, \dots, c_{2k})$ of a k-slow walk, the situation is reversed: The only index $1 < j < 2k$ with c_j being in between c_{j-1} and c_{j+1} is the index $j = k$.

In order to apply the lexicographic and reversed thinning recursively, we have to change the coloring of the subgraph. Let $G = (A, B, E)$ be a bipartite graph with a proper edge coloring given by $\chi : E \to \{1, \dots, d\}$. Let $G' = (A, B, E')$ be the result of the lexicographic or the reversed thinning of G. Recall that the edges in E' have a type $2 \leq i \leq t$ with $t = \lceil (\log d)/2 \rceil + 1$. The *type edge coloring* of G' is the map $\chi' : E' \to \{1, \dots, t-1\}$ defined by $\chi'(e) = t + 1 - i$ for an edge $e \in E'$ of type i. By Lemma 2.8(a), χ' is a proper edge coloring of G'.

Lemma 2.12. *Let $G = (A, B, E)$ be a bipartite graph with proper edge coloring given by $\chi : E \to \{1, \dots, d\}$. Let $G' = (A, B, E')$ be the result of the lexicographic thinning of G. Let χ' be the type edge coloring of G' and let $k \geq 2$. If a subgraph $G'' = (A, B, E'')$ of G', with its edge coloring given by χ', has no (k', A)-slow walk for $2 \leq k' < k$, then G'', with its edge coloring given by χ, has no (k, B)-slow walk.*

Notice that the $k = 2$ case of this lemma gives a second proof of Lemma 2.9.

Proof. Let $W = v_0 v_1 \dots v_{2k}$ be a walk in G'' starting at $v_0 \in B$, and let its χ-coloring be $(c_1, c_2, \dots, c_{2k})$. Assume that $c_i > c_{i+1}$ or $c_i < c_{i+1}$ for $1 \leq i < 2k$ as required in the definition of a (k, B)-slow walk. We need to show $c_1 < c_{2k}$.

We identify the colors of χ with the triplets $(a, i, z) \in H$ as in the definition of lexicographic thinning. We let $c_j = (a_j, i_j, z_j)$. The χ'-coloring of W is $(t + 1 - i_1, \dots, t + 1 - i_{2k})$. We have $i_j \neq i_{j+1}$ for $1 \leq j < 2k$.

For $1 \leq j < k$, the colors c_{2j} and c_{2j+1} are colors of distinct edges incident to $v_{2j} \in B$, and so by Lemma 2.8(b), their class is the same: $a_{2j} = a_{2j+1}$. We consider lexicographic thinning, and so the order between c_{2j} and c_{2j+1} is the same as the order between their types: i_{2j} and i_{2j+1}. For $1 \le j < k/2$, we have $i_{2j} < i_{2j+1}$, but for $k/2 \le j < k$, we have $i_{2j} > i_{2j+1}$.

For $1 \le j \le k$, the colors c_{2j-1} and c_{2j} are colors of edges incident to $v_{2j-1} \in A$. By Lemma 2.8(c), the classes of these colors do not agree, and the orderings between the classes, between the types, and between the colors themselves are the same. Thus, for $1 \le j \le k/2$, we have $a_{2j-1} > a_{2j}$ and $i_{2j-1} > i_{2j}$. For $k/2 < j \le k$, we have $a_{2j-1} < a_{2j}$ and $i_{2j-1} < i_{2j}$. Also, by Lemma 2.8(c) for all $1 \le j \le k$, we have $P(a_{2j-1}, a_{2j}) = \min(i_{2j-1}, i_{2j})$.

The sequence $a_1, a_2, \ldots, a_k$ is monotone decreasing and it changes only in every other step. The first positions of change between distinct consecutive elements are $i_2, i_4, \ldots, i_{2\lfloor k/2 \rfloor}$. So we have $a_1 > a_k$ and $P(a_1, a_k) = \min(S_1)$ for the set $S_1 = \{i_2, i_4, \ldots, i_{2\lfloor k/2 \rfloor}\}$.

Similarly, $a_k, a_{k+1}, \ldots, a_{2k}$ is monotone increasing and it changes only in every other step. The first positions of change between distinct consecutive elements are $i_{2\lfloor k/2 \rfloor+1}, \ldots, i_{2k-3}, i_{2k-1}$. So we have $a_k < a_{2k}$ and $P(a_k, a_{2k}) = \min(S_2)$ for the set $S_2 = \{i_{2\lfloor k/2 \rfloor+1}, \ldots, i_{2k-3}, i_{2k-1}\}$.

Let's consider an arbitrary value $1 \le j < k/2$ and let $2 \le k' = k - 2j + 1 < k$. Consider the $2k'$ long middle portion W' of the walk W: Let $W' = v_{2j-1} v_{2j} \ldots v_{2k-2j+1}$. This is a walk of length $2k'$ in G'' starting at $v_{2j-1} \in A$. By our assumption on G'', this is not a (k', A)-slow walk if considered with the coloring χ'. But the χ'-coloring of W' is $(t+1-i_{2j}, t+1-i_{2j+1}, \ldots, t+1-i_{2k-2j+1})$, and the consecutive values in this list compare as required for a (k', A)-slow walk. Therefore, we must have $t+1-i_{2j} < t+1-i_{2k-2j+1}$.

We have just proved $i_{2j} > i_{2k-2j+1}$ for $1 \le j < k/2$. For even k and $j = k/2$, the same formula compares the types of two consecutive edges of W, and we have already seen its validity in that case too. For every element of the set S_1, we have just found a smaller element of the set S_2. Therefore, $\min(S_1) > \min(S_2)$. Using that $a_{2k} > a_k$ and $P(a_1, a_k) = \min(S_1) > \min(S_2) = P(a_k, a_{2k})$, Lemma 2.7 gives $a_1 < a_{2k}$. This implies $c_1 < c_{2k}$ and finishes the proof of the lemma. □

Lemma 2.13. *Let $G = (A, B, E)$ be a bipartite graph with proper edge coloring given by $\chi : E \to \{1, \ldots, d\}$. Let $G' = (A, B, E')$ be the result of the reversed thinning of G. Let χ' be the type edge coloring of G', and let $k \ge 2$. If a subgraph $G'' = (A, B, E'')$ of G', with its edge coloring given by χ', has no (k', A)-slow walk for $2 \le k' < k$, then G'', with its edge coloring given by χ, has no k-fast walk.*

Notice that the $k = 2$ case of this lemma gives a second proof of Lemma 2.10.

Proof. The proof of this lemma is very similar to that of Lemma 2.12.

Let $W = v_0 v_1 \ldots v_{2k}$ be a walk in G'', and let its χ-coloring be $(c_1, c_2, \ldots, c_{2k})$. Assume that $c_1 > c_2 > \cdots > c_k < c_{k+1} < c_{k+2} < \cdots < c_{2k}$, as required in the definition of a k-fast walk. We need to show $c_1 < c_{2k}$. Instead, we prove the slightly stronger statement that the class of c_1 is smaller than the class of c_{2k}. We first do that for walks starting in B: Assume that $v_0 \in B$.

We identify the colors of χ with the triplets $(a,i,z) \in H$ as in the definition of reverse thinning. We let $c_j = (a_j, i_j, z_j)$. Note that the χ'-coloring of W is $(t+1-i_1, \ldots, t+1-i_{2k})$. We have $i_j \neq i_{j+1}$ for $1 \leq j < 2k$.

For $1 \leq j < k$, the colors c_{2j} and c_{2j+1} are colors of distinct edges incident to $v_{2j} \in B$, so by Lemma 2.8(b) their classes are the same: $a_{2j} = a_{2j+1}$. Thus, the order between c_{2j} and c_{2j+1} is determined by the order between i_{2j} and i_{2j+1}, but as we use the reversed ordering in H, the order between c_{2j} and c_{2j+1} is reversed compared to the order between i_{2j} and i_{2j+1}. Specifically, for $1 \leq j < k/2$, we have $i_{2j} < i_{2j+1}$, and for $k/2 \leq j < k$, we have $i_{2j} > i_{2j+1}$. For $1 \leq j \leq k$, the colors c_{2j-1} and c_{2j} are colors of edges incident to $v_{2j-1} \in A$. By Lemma 2.8(c), the classes of these colors do not agree, and the orderings between the classes, between the types, and between the colors themselves are the same. Thus, for $1 \leq j \leq k/2$, we have $a_{2j-1} > a_{2j}$ and $i_{2j-1} > i_{2j}$. For $k/2 < j \leq k$, we have $a_{2j-1} < a_{2j}$ and $i_{2j-1} < i_{2j}$. Also, by Lemma 2.8(c), for $1 \leq j \leq k$, we have $P(a_{2j-1}, a_{2j}) = \min(i_{2j-1}, i_{2j})$.

At this point we have the same ordering of the classes and types of the coloring of W as in the proof of Lemma 2.12. We also have the same assumption that G'', with the edge coloring χ', has no (k', A)-slow walk for $2 \leq k' < k$. So we arrive at the same conclusion, $a_1 < a_{2k}$, with an identical proof.

We finish the proof of the lemma by considering the alternative case when W starts in A. Now c_1 and c_2 are colors of edges sharing a vertex $v_1 \in B$, and so by Lemma 2.8(b) their classes are equal. Similarly, the classes of c_{2k-1} and c_{2k} are equal, so it is enough to prove that the class of c_2 is smaller than the class of c_{2k-1}. For $k = 2$, this follows directly from Lemma 2.8(c). For $k > 2$, the walk $W' = v_1 v_2 \ldots v_{2k-1}$ is exactly the type of walk we considered for $k_0 = k-1$. As it starts in B, we have already proved that the class of its first edge is smaller than the class of its last edge. This finishes the proof of the case of a walk starting in A and also the proof of Lemma 2.13. □

Lemmas 2.12 and 2.13 set the stage to use the thinning procedures recursively to get subgraphs avoiding (k,B)-slow or k-fast walks. In a single application of either thinning procedure, the number d of colors in the original coloring is replaced by $t - 1 = \lceil \log d/2 \rceil$ colors in the type coloring. Here $4(t-1) > \log(4d)$, so after k recursive calls, we still have more than $\log^{(k)}(4d)/4$ colors, where $\log^{(k)}$ stands for the k-times iterated log functions. [In the degenerate case where the number of colors is decreased to 1, the graph is a matching and no further thinning is required.] Making optimal random choices, we may assume that each thinning procedure yields at least the expected number of edges. Thus, the ratio of the number of edges and the number of colors decreases by at most a factor of 240 in each iteration. Clearly, the only interesting case is when the original average degree was $\Theta(d)$, in which case the average degree after k iterations remains $\Theta(\log^{(k)} d)$. The constant of proportionality depends on k.

Theorem 2. *Let $G = (A,B,E)$ be a bipartite graph with a d-edge coloring and let $k \geq 2$. There exists a subgraph $G' = (A,B,E')$ of G without a (k',B)-slow walk for* *any* $2 \le k' \le k$ *and with* $|E'| > \frac{\log^{(k-1)} d}{4 \cdot 240^{k-1} d}|E|$*. Similarly, there exists a subgraph* $G'' = (A, B, E'')$ *of* G *without a* k'*-fast walk for any* $2 \le k' \le k$ *and with* $|E''| > \frac{\log^{(k-1)} d}{4 \cdot 240^{k-1} d}|E|$.

Proof. We apply the thinning procedures recursively. First, we use lexicographic and reversed thinning to obtain subgraphs G_1 and G_2 of G, respectively. We make sure these graphs have at least as many edges as the expected number given in Lemma 2.11. If $k = 2$, we are done; $G' = G_1$ and $G'' = G_2$ satisfy the conditions of the theorem. Otherwise, we consider G_1 and G_2 with the type edge coloring. We find recursively their subgraphs G' and G'', respectively, avoiding (k', A)-slow walks for $2 \le k' \le k - 1$. This can be done because the sides A and B play symmetric roles. Finally, we apply Lemmas 2.12 and 2.13 to see that the subgraphs G' and G'', if considered with the original edge coloring of G, avoid all walks required in the theorem. The number of edges guaranteed in the subgraphs is calculated in the paragraph preceding the theorem and is at least the stated bound. □

2.4 k-Flat Graphs

In this subsection, we establish that by removing a linear number of edges from a k-fast walk free graph, the resulting graph has special structural properties. We note here that the recursive thinning construction that we used to arrive at k-fast walk-free graphs results in a graph that itself is k-flat as defined below. We chose, however, to keep the inductive part of the proof simple and concentrated only on (k, B)-slow and k-fast walks. We derive the more complicated properties from these simpler ones. Note that in this subsection we do not use that our graphs are bipartite.

Let G be a graph and χ a proper edge coloring of G. We define the *shaving* of the graph G to be the subgraph obtained from G by deleting the edge with the largest color incident to every (nonisolated) vertex. Clearly, we delete at most n edges, where n is the number of vertices in G. We define the k-*shaving* of G to be the subgraph obtained from G by repeating the shaving operation k times. Clearly, we delete at most kn edges for a k-shaving.

Let W be a walk of length m in a properly edge-colored graph G, and assume its coloring is $(c_1, \ldots, c_m)$. We define the *height function* h_W from $\{1, \ldots, m\}$ to the integers recursively, letting $h_W(1) = 0$ and

$$h_W(i+1) = \begin{cases} h_W(i) + 1 & \text{if } c_{i+1} > c_i \\ h_W(i) - 1 & \text{if } c_{i+1} < c_i \end{cases}$$

for $1 \le i < m$. Note that $h_W(i) + i$ is always odd. This function considers how the colors of the edges in the walk change, in particular, how many times the next color is larger and how many times it is smaller than the previous color.

We call a graph G with proper edge coloring *k-flat* if the following is true for every walk W in G. Let $m \ge 2$ be the length of W, let $(c_1, \ldots, c_m)$ be the coloring of W, and assume that the height function satisfies $h_W(i) < 0$ for $2 \le i \le m$. If $m \le 2k+1$ or $h_W(i) \ge -k$ for all i, then $c_1 > c_m$.

Lemma 2.14. *Let G be a properly edge-colored graph. Let $k \ge 1$ and assume G has no k'-fast walk for $2 \le k' \le k$. Then the $(k-1)$-shaving G' of G is k-flat.*

Proof. We prove the following slightly stronger statement by induction on m. Let $W = v_0 \dots v_m$ be a walk of length m in G with coloring $(c_1, \dots, c_m)$. Let $1 \le j \le m$ be the largest index such that $h_W(j) = 1 - j$. Assume the walk $v_j v_{j+1} \dots v_m$ is in the $(k-1)$-shaving G' of G. Also, assume that $h_W(i) < 0$ for $2 \le i \le m$. If $m \le 2k+1$ or $h_W(i) \ge -k$ for all i, then we claim $c_1 > c_m$. This statement is stronger than Lemma 2.14 since it allows for the initial decreasing segment of W to be outside G'.

If $j = m$, the statement of the claim is obvious from the definition of the height function. This covers the $m = 2$ and $m = 3$ base cases. Let $m \ge 4$, and assume the statement is true for walks of length $m-1$ and $m-2$.

If $j > k+1$, we have $h_W(j) = 1 - j < -k$, and so we must have $m \le 2k+1$. Consider the walk $W' = v_1 \dots v_m$ of length $m-1$. We have $h_{W'}(i) = 1 - i < 0$ for $2 \le i < j$ and $h_{W'}(i) \le h_{W'}(j-1) + (i - (j-1)) = 3 + i - 2j < 0$ for $j \le i \le m-1$. Thus, the inductive hypothesis is applicable, and we get $c_1 > c_2 > c_m$.

Finally, consider the $j \le k+1$ case. As the trivial $j = m$ case was already dealt with, we also assume $j < m$. Clearly, $h_W(2) < 0$ implies $j \ge 2$. We chose a $w_0 \dots w_{j-2}$ walk in G ending at $w_{j-2} = v_j$ and with coloring $(c'_1, \dots, c'_{j-2})$ satisfying $c'_1 > c'_2 > \dots > c'_{j-2} > c_{j+1}$. This is possible since the edge $v_j v_{j+1}$ is in the $(k-1)$-shaving G' of G, so we can find the edge $w_{j-3} w_{j-2}$ in the $(k-2)$-shaving of G, $w_{j-4} w_{j-3}$ in the $(k-3)$-shaving, and so on. We must have $c_1 > c'_1$, as otherwise $w_0 w_1 \dots w_{j-3} v_j v_{j-1} \dots v_0$ is a $(j-1)$-fast walk, and no such walk exists in G. Now consider the walk $W' = w_0 w_1 \dots w_{j-3} v_j v_{j+1} \dots v_m$. This is a walk of length $m-2$ and satisfies $h_{W'}(i) = 1 - i$ for $1 \le i \le j-1$ and $h_{W'}(i) = h_W(i+2)$ for $j-1 \le i \le m-2$. All requirements of the inductive hypothesis are satisfied, so we have $c'_1 > c_m$. Thus, $c_1 > c'_1 > c_m$, as claimed. □

Corollary 2.1. *Let $G = (A, B, E)$ be a bipartite graph with a d-edge coloring and let $k \ge 2$. There exists a k-flat subgraph $G' = (A, B, E')$ of G with $|E'| > \frac{\log^{(k-1)} d}{4 \cdot 240^{k-1} d} |E| - (k-1)(|A| + |B|)$.*

Proof. Combine Theorem 2 with Lemma 2.14 and the fact that $(k-1)$-shaving keeps all but at most $(k-1)(|A| + |B|)$ edges of G. □

The final lemma in this section is a simple but useful observation on k-flat graphs. It can also be stated for longer walks with height function bounded from below, but for simplicity we restrict attention to short walks.

Lemma 2.15. *Let $k \ge 1$ and let G be a properly edge-colored k-flat graph. Let $W = v_0 \dots v_m$ be a walk in G of length $m \le 2k+1$ with coloring $(c_1, \dots, c_m)$. If $c_1 \ge c_i$ for all $1 \le i \le m$, then $h_W(i) \le 0$ for all $1 \le i \le m$.*

Proof. We prove the contrapositive statement. Assume $h_W(i) > 0$ for some $1 \leq i \leq m$, and let i_0 be the smallest such index. Clearly, $i_0 \geq 2$, $h_W(i_0) = 1$, and for the walk $W' = v_{i_0} v_{i_0-1} \dots v_0$, we have $h_{W'}(i) = h_W(i_0 - i + 1) - 1 < 0$ for $1 < i \leq i_0$. So by the definition of k-flatness, we have $c_{i_0} > c_1$. □

3 Locally Plane Graphs

Locally plane graphs were introduced in the paper [7] (though the name appears first in this chapter). That paper gives a simple construction for 3-locally plane graphs. We recall (a simplified version of) the construction, as it is our starting point.

3.1 Construction of 3*-Locally Plane Graphs in [7]*

Let $d \geq 1$ and consider the orthogonal projection of (the edge graph of) the d-dimensional hypercube into the plane. A suitable projection of the "middle layer" of the hypercube provides the 3-locally plane graph. Here is the construction in detail:

Let $d \geq 1$ be fixed and set $b = \lfloor d/2 \rfloor$. The bit at position i in $x \in \{0,1\}^d$ (the ith coordinate) is denoted by x_i for $1 \leq i \leq d$. We let $A = \{x \in \{0,1\}^d \mid \sum_{i=1}^d x_i = b\}$ and $B = \{x \in \{0,1\}^d \mid \sum_{i=1}^d x_i = b+1\}$. The abstract graph underlying the geometric graph to be constructed is $G_d = (A, B, E)$ with $(x,y) \in E$ if $x \in A$, $y \in B$, and x differs from y in a single position. This is the middle layer of the d-dimensional hypercube. We define the edge coloring $\chi : E \to \{1, \dots, d\}$ that colors an edge $e = (x,y) \in E$ by the unique position $\chi(e) = i$ with $x_i \neq y_i$. Notice that this is a proper edge coloring. The number of vertices is $n = |A| + |B| = \binom{d}{b} + \binom{d}{b+1} \leq 2^d$, and the number of edges is $|E| = \binom{d}{b}(d-b) > nd/4$. The average degree is greater than $d/2 \geq \log n/2$.

To make the abstract graph G_d into a geometric graph, we project the hypercube into the plane. We give two possible projections here. The first is more intuitive and is closer to the actual construction in [7]. We let $a_i = (10^i, i \cdot 10^i)$ for $1 \leq i \leq d$ and use this vector as the projection of the edges of color i. We will use that among the vectors a_i, a higher index means a higher slope and much greater length. We identify the vertex $x \in A \cup B$ with the point $P_x = \sum_{i=1}^d x_i a_i$. The edges are represented by the straight-line segment connecting their endpoints.

We give the second construction to obtain a graph where all edges are very close in length and direction. Let $0 < \varepsilon < 10^{-d}$ be arbitrary, consider the vectors $b_i = (1 + 10^i \varepsilon, \varepsilon^{d+1-i}(1 + 10^i \varepsilon))$, and identify a point $x \in A \cup B$ with $Q_x = \sum_{i=1}^d x_i b_i$.

It is easy to verify that we get a geometric graph in both cases (i.e., the vertices are mapped to distinct points and no edge passes through a vertex that is not its endpoint). Note that edges of color i are all translates of the same vector a_i or b_i. We do not introduce separate notations for the two geometric graphs constructed this way, as they will only be treated separately in the proof of Lemma 3.1, where we refer to them as the first and second realizations of G_d.

3.2 Self-Intersecting Paths in G_d

In [7], a graph very similar to G_d was shown to be 3-locally plane. Here we do more; we analyze all self-intersecting paths of G_d as follows.

Lemma 3.1. *Let W be a walk in G_d with coloring $(c_1,\dots,c_m)$ satisfying $c_1 \geq c_m$. Assume that W and all its nonempty subwalks have a unique edge of maximal color. The first and last edges of W cross in either geometric realization if and only if m is even and there is an odd index $1 < j < m$ satisfying $c_1 > c_m > c_j \geq c_i$ for all $1 < i < m$.*

Proof. Let $W = v_0 \dots v_m$. Note that the first and last edges cross if and only if v_0 and v_1 are on different sides of the line ℓ through v_{m-1} and v_m and, similarly, v_{m-1} and v_m are on different sides of the line ℓ' through v_0 and v_1. To analyze such separations, consider the projection π_i to the y-axis parallel to edges of color i.

Let's consider the first realization of G_d with the vectors a_i. We have $\pi_i(x,y) = y - ix$, and the projection of the vector a_j is of length $|i-j|10^j$. Thus, higher-colored edges map to longer intervals (except color i itself). Under the projection π_{c_1}, the direction of the highest colored edge in the walks $v_1 \dots v_{m-1}$ (respectively, $v_1 \dots v_m$) determines which side of the line ℓ' the point v_{m-1} (respectively, v_m) lies. Indeed, this highest color cannot be c_1, so the projections of the other edges will be much shorter, and by the unique maximal color property, we see that the walk contains at most 2^{k-1} edges having the kth largest color, so these shorter projections cannot add up to be more than the projection of the largest edge.

We can only have v_{m-1} and v_m lying on different sides of ℓ' if these edges of maximal color are distinct; thus, we must have $c_m > c_i$ for all $1 < i < m$. From $c_1 \geq c_m$ and the unique maximal color property, we have $c_1 > c_m$. Taking c_j to maximize c_i for $1 < i < m$ (this is unique again), we have $c_1 > c_m > c_j \geq c_i$ for all $1 < i < m$.

It is left to prove that the first and last edges of W cross if and only if m is even and j is odd.

To prove this claim, one has to use that G_d is bipartite with vertex sets A and B, and every edge of color c is a translate of the vector a_c with its head in B and tail in A. Thus, the vector $v_{i-1}v_i$ is either a_{c_i} or $-a_{c_i}$ depending on the parity of i. Which sides of ℓ', v_{m-1}, and v_m lie is determined by the projections of the jth and last edge, so they are on opposite sides if j and m have different parities. Similarly, the sides of ℓ on which v_0 and v_1 lies are determined by the π_{c_m} projection of the first and jth edges, but as we have $c_1 > c_m > c_j$, they are on different sides if 1 and j have the same parity. See Fig. 1 for a rough depiction of all four cases. This finishes the proof of the claim and the part of the lemma regarding the first realization of G_d.

The proof for the second realization of G_d as a geometric graph (involving the vectors b_i) is slightly more complicated. We have $\pi_i(x,y) = y - \varepsilon^{d+1-i}x$ and $\pi_i(b_j) = \varepsilon^{d+1-j}(1+10^j\varepsilon) - 10^j\varepsilon^{d+2-i} - \varepsilon^{d+1-i}$. The ε^{d+1-i}-terms alternate in sign in the projection on the edges and cancel completely for a walk of even length. For a walk of odd length, a single such term remains, but it is dominated by the other

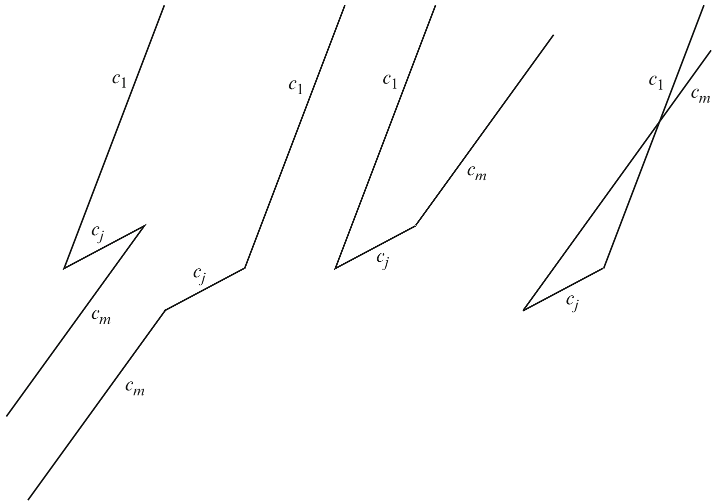


Fig. 1 Rough picture of W in the first realization of G_d in the four cases according to the parities of m and j. The scale is set to match the edge $v_{j-1}v_j$ of color c_j, shorter edges are approximated by zero, and only initial segments of the two longer edges are depicted

terms if the walk has an edge with color above i. If, however, no such edge exists, the remaining uncanceled ε^{d+1-i}-term dominates the other terms in the projection. Thus, if the projection of the unique edge with the largest color of a walk has color $c > i$ or if $c < i$ but the walk has even length, then the sign of the π_i projection of the edge with the largest color determines the sign of the projection of the entire walk. But in case $c < i$ and the walk has odd length, then the noncancelling ε^{d+1-i}-term determines the sign.

Let the edge $v_{j-1}v_j$ be the one with the unique largest color in the walk $v_1 \dots v_{m-1}$. A case analysis of the parities of j, m and whether $c_j > c_1$ or $c_j > c_m$ hold shows that the first and last edges of W cross if and only if $c_m > c_j$, j is odd, and m is even—as claimed in the lemma. □

We call $v_m \dots v_0$ the *reverse* of the walk $W = v_0 \dots v_m$.

Lemma 3.2. *Let $k \geq 1$ and let G' be a k-flat subgraph of G_d. If the length of a walk W in G' does not exceed $2k+1$, then W has unique edge of largest color.*

Proof. Let W be a walk of length $3 \leq m \leq 2k+1$ in G' with coloring $(c_1, \dots, c_m)$ and assume the largest color is not unique. We may assume $c_1 = c_m > c_i$ holds for all $1 < i < m$; otherwise, one can take a suitable subwalk of W. By Lemma 2.15, we have $h_W(m) \leq 0$. Consider W' the reverse of the walk W. Clearly, $h_{W'}(m) = -h_W(m)$, but also by Lemma 2.15, we have $h_{W'}(m) \leq 0$. So we must have $h_W(m) = 0$ and m must be odd.

The contradiction that proves the statement of the lemma comes from the simple observation that between two consecutive appearances of a color in any walk of G_d, there always are an even number of edges, so m must be even. To see this, recall that G_d is bipartite with sides A and B, where A consists of the 0–1 sequences of length d with $\lfloor d/2 \rfloor$ ones, while the 0–1 sequences in B contain one more ones. The A end of an edge of color c has 0 at position c, while the B end of this edge has 1 there. Along edges of other colors, bit c does not change. Thus, if a walk traverses an edge of color c from A to B say, then along the walk, bit c remains 1 until the next time the walk traverses an edge of color c, and this has to be from B to A. □

We note that Lemma 3.2 immediately implies that the girth of a k-flat subgraph of G_d is at least $2k+2$. This estimate can be improved by observing that any cycle in G_d has an even number of edges of any color (one has to flip a bit even times to get beck to the original state). In particular, any cycle has at least two occurrences of the largest color. These edges break the cycle into two paths sharing two edges. Both paths have to be of length at least $2k+2$; the length of the cycle is at least $4k+2$.

As every properly edge-colored graph is 1-flat, the $k=1$ case of the next lemma establishes that G_d is 3-locally plane.

Lemma 3.3. *For any $k \geq 1$, any k-flat subgraph of G_d is $(2k+1)$-locally plane in either realization.*

Proof. Let G' be a k-flat subgraph of G_d. We need to show that no walk (or path) W of length $m \leq 2k+1$ is self-intersecting. It is clearly enough to show that the first and last edges of W do not cross, and we may assume that the color of the first edge is not smaller than that of the last edge (otherwise, simply consider the same walk reversed). By Lemma 3.2, W and all its subwalks have a unique edge of maximal color, so Lemma 3.1 applies. It is enough to show that the coloring $(c_1, \ldots, c_m)$ of W does not satisfy the conditions of Lemma 3.1. Assume the contrary. So m is even, and there is an odd index j such that $c_1 > c_m > c_j \geq c_i$ for all $1 < i < m$.

Consider the walk $W_1 = v_m v_{m-1} \ldots v_{j-1}$ of length $m-j+1$. Its coloring is $(c_m, \ldots, c_j)$ and c_m is its largest color. By Lemma 2.15, $h_{W_1}(m-j+1) \leq 0$. In fact, as $m-j+1$ is even, $h_{W_1}(m-j+1)$ is odd, so we have $h_{W_1}(m-j+1) \leq -1$.

Consider the walk $W_2 = v_j v_{j-1} \ldots v_1$ of length $j-1$ and its coloring $(c_j, \ldots, c_2)$. The largest color in W_2 is c_j, so we have $h_{W_2}(j-1) \leq 0$ by Lemma 2.15. And again, by parity considerations, $h_{W_2}(j-1) \leq -1$.

It is easy to see that $h_W(m) = -1 - h_{W_2}(j-1) - h_{W_1}(m-j+1)$. So we have $h_W(m) \geq -1+1+1 = 1$, contradicting Lemma 2.15. The contradiction proves Lemma 3.3. □

Theorem 3. *For any fixed $k > 0$ and large enough n, there exists a $(2k+1)$-locally plane graph on n vertices having at least $(\frac{\log^{(k)} n}{240^k} - k)n$ edges. Given two arbitrary disks in the plane, one can further assume that all edges of these graph connect a vertex inside one disk to one inside the other disk.*

Proof. Simply combine the results of Corollary 2.1 and Lemma 3.3. If n is not the size of the vertex set of G_d for any d, add isolated vertices to the largest G_d with fewer than n vertices. Use the second realization of G_d with a small enough $\varepsilon > 0$ to obtain a geometric graph with all edges connecting two small disks and apply a homothety and a rotation to get to the desired disks. □

4 Discussion on Optimality of Thinning

The maximum number of edges of a 3-locally plane graph on n vertices is $\Theta(n \log n)$, as proved in [7]. The lower bound is reproduced here by the $k = 1$ case of Theorem 3, which is therefore tight. The upper bound of [7] extends to *x-monotone topological graphs*, i.e., when the edges are represented by curves with the property that every line parallel to the y-axis intersects an edge at most once. Without this artificial assumption on x-monotonicity, only much weaker upper bounds are known. For higher values of k, we do not have tight results even if the edges are straight-line segments as considered in this chapter. While the number of edges in a k-locally plane graph constructed here deteriorates very rapidly with the increase of k, the upper bound hardly changes. In fact, the only known upper bound better than $O(n \log n)$ is for 5-locally plane graphs: They have $O(n \log n / \log \log n)$ edges, as shown in [7]. Slightly better bounds are known for geometric (or x-monotone topological) graphs with the additional condition that a vertical line intersects every edge. If such a graph is $(2k)$-locally plane for $k \geq 2$, then it has $O(n \log^{1/k} n)$ edges. The first realization of G_d does not satisfy this condition, but the second one does. Still, the lower and upper bounds for this restricted problem are far apart: For 4-locally plane graphs with a cutting line, the upper bound on the number of edges is $O(n\sqrt{\log n})$, while the construction gives $\Omega(n \log \log n)$.

Although we cannot establish that the locally plane graphs constructed are optimal, we can prove that the thinning procedure we use is optimal within a constant factor. It follows that any 4-locally plane subgraph of either realization of G_d has $O(n \log \log n)$ edges. This optimality result below refers to a single step of the thinning procedure. It would be interesting to establish a strong upper bound on the number of edges of a k-flat graph for $k \geq 3$.

Let us mention here that the thinning procedures described in this chapter found application in the extremal theory of 0-1 matrices; see [12], where the result is shown to be optimal within a constant factor. Consider an n by n 0-1 matrix that has no 2 by 3 submatrix of either of the following two forms:

$$\begin{pmatrix} 1 & 1 & * \\ 1 & * & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & * & 1 \\ * & 1 & 1 \end{pmatrix},$$

where the $*$ can represent any entry. The maximal number of 1 entries in such a matrix is $\Theta(n \log \log n)$, as proved in [12]. The construction proving the lower bound is based on lexicographic thinning.

The following lemma shows that the number of edges in the subgraphs claimed in Lemma 2.6 and Theorem 1 is optimal in a very strong sense: No properly edge-colored graph with significantly more edges than the ones guaranteed by the above results can avoid heavy paths (slow or fast walks, respectively).

Lemma 4.1. *Let $G=(A,B,E)$ be a bipartite graph with proper edge coloring given by $\chi : E \to \{1,\dots,d\}$.*

a) If G does not have a heavy path, then $|E| \le 2\sqrt{d|A|\,|B|} \le (|A|+|B|)\sqrt{d}$.
b) If G does not have a slow walk, then $|E| \le (|A|+|B|)(\log d+2)$.
c) If G does not have a fast walk, then $|E| \le 2(|A|+|B|)(\log d+2)$.

Proof. For any vertex $z \in A \cup B$, denote by $m(z)=\max(\chi(e))$, where the maximum is for edges e incident to z. For an edge $e=(x,y) \in E$, we define its *weight* to be $w(e)=m(x)-\chi(e)$, while its *B-weight* is $w_B(e)=m(y)-\chi(e)$. Clearly, both $w(e)$ and $w_B(e)$ are integers in $[0,d-1]$.

To prove part (a) of the lemma, assume G does not contain a heavy path. We set a threshold parameter $t=\lfloor\sqrt{d|A|/|B|}\rfloor$ and call an edge e *B-light* if $w_B(e)<t$; otherwise, e is *B-heavy*.

All the edges incident to a vertex $y \in B$ have different colors; thus, they also have different B-weights, so at most t of them can be B-light. The total number of B-light edges is at most $|B|t \le \sqrt{d|A|\,|B|}$.

Now consider two edges $e_1=(x,y_1)$ and $e_2=(x,y_2)$ incident to vertex $x \in A$. Assume $\chi(e_2) \le \chi(e_1)$. If $w_B(e_2)>0$, we can extend the path formed by these two edges with the edge e_3 incident to y_2 having maximum color $\chi(e_3)=m(y_2)$. Clearly, $\chi(e_3)=\chi(e_2)+w_B(e_2)$, and the resulting path is heavy unless $\chi(e_2)>\chi(e_1)+w(e_2)$. Therefore, the number of B-heavy edges incident to x is at most $d/(t+1)$. The total number of B-heavy edges is at most $|A|d/(t+1) \le \sqrt{d|A|\,|B|}$.

For the total number of edges, we have $|E| \le 2\sqrt{d|A|\,|B|} \le (|A|+|B|)\sqrt{d}$.

For parts (b) and (c) of the lemma, consider an edge $e=(x,y) \in E$. If $w(e)=0$, we call the edge e *maximal*. Clearly, there are at most $|A|$ maximal edges. If e is not maximal, we define $n(e)$ to be the "next-larger colored edge at x;" i.e., $n(e)$ is the edge in E having minimal color $\chi(n(e))$ among edges incident to x and satisfying $\chi(n(e))>\chi(e)$. We define the *gap* of e to be $g(e)=\chi(n(e))-\chi(e)$. Clearly, $0<g(e) \le w(e)$. We call the edge e *heavy* if $w(e)>2g(e)$; otherwise, e is *light*. Recall that for maximal edges, $n(e)$ and $g(e)$ are not defined and maximal edges are neither light nor heavy.

Let e_1 and e_2 be distinct edges in E incident to a vertex $x \in A$. If $\chi(e_1)<\chi(e_2)$, then $w(e_1) \ge w(e_2)+g(e_1)$. If e_1 is light, $w(e_1) \ge 2w(e_2)$ follows. Therefore, at most $\lceil\log d\rceil$ light edges are incident to $x \in A$. Thus, the total number of light edges in E is at most $\lceil\log d\rceil|A|$.

For part (b) of the lemma, assume G does not contain a slow walk. Let $e_2=(x_2,y)$ and $e_3=(x_3,y)$ be distinct nonmaximal edges in E and assume $\chi(e_2)<\chi(e_3)$. Let $e_1=n(e_2)$, and let e_4 be the maximal edge incident to x_3 in G. We have $\chi(e_1)=\chi(e_2)+g(e_2)$ and $\chi(e_4)=\chi(e_3)+w(e_3)$. The edges e_1, e_2, e_3, and e_4 cannot form a slow walk. As $\chi(e_1)>\chi(e_2)<\chi(e_3)<\chi(e_4)$, we must have $\chi(e_4)<\chi(e_1)$.

This implies $g(e_2) > w(e_3)$, and if e_2 is heavy, $w(e_2) > 2w(e_3)$. Therefore, at most $\lceil \log d \rceil$ heavy edges can be incident to $y \in B$. The total number of heavy edges in G is at most $(\lceil \log d \rceil |B|$.

For the total number of edges, we add the bound obtained for light, heavy, and maximal edges and get $|E| \leq (|A| + |B|)\lceil \log d \rceil + |A|$.

Finally, for part (c) of the lemma, we assume G does not contain a fast walk. Let E_1 consist of the edges $(x,y) \in E$ for which $m(x) \geq m(y)$. Assume without loss of generality that $|E_1| \geq |E|/2$. If this is not the case, consider the same graph with its sides switched. We use here that the definition of a fast walk and the claimed bound on the number of edges are both symmetric in the color classes.

As in the previous case, consider two nonmaximal edges $e_2 = (x_2,y)$ and $e_3 = (x_3,y)$ in E_1 with $\chi(e_2) < \chi(e_3)$. Let e_1 be the maximal edge incident to x_2 and let $e_4 = n(e_3)$. As $\chi(e_2) < \chi(e_3) < \chi(e_4)$ but G does not contain a fast walk, we must have $\chi(e_1) < \chi(e_4)$. As $e_2 \in E_1$, we must also have $\chi(e_1) = m(x_2) \geq m(y) \geq \chi(e_3)$. If e_3 is heavy, we also have

$$\begin{aligned} \chi(e_3) + w(e_3) - m(y) &> \chi(e_3) + 2g(e_3) - m(y) \\ &\geq 2(\chi(e_3) + g(e_3) - m(y)) \\ &= 2(\chi(e_4) - m(y)) \\ &> 2(\chi(e_1) - m(y)) \\ &= 2(\chi(e_2) + w(e_2) - m(y)). \end{aligned}$$

For all the heavy edges $e \in E_1$ incident to $y \in B$, the values $\chi(e) + w(e) - m(y)$ increase strictly more than by a factor of 2. As these values are integers from $[0, d-1]$, there are at most $\lceil \log d \rceil$ heavy edges in E_1 incident to y. The total number of heavy edges in E_1 is at most $\lceil \log d \rceil |B|$.

For the total number of edges in E_1, we sum our bound on heavy edges in E_1 and the bounds on the light and maximal edges in E. We obtain $|E_1| \leq (|A| + |B|)\lceil \log d \rceil + |A|$. Finally, we get $|E| \leq 2(|A| + |B|)\lceil \log d \rceil + 2|A|$. □

Acknowledgements The author's research was partially supported by NSERC Grant 329527 and by OTKA Grants T-046234, AT048826, and NK-62321.

References

1. E. Ackerman, On the maximum number of edges in topological graphs with no four pairwise crossing edges. Discrete Comput. Geom. **41**(3), 365–375 (2009)
2. P. Agarwal, B. Aronov, J. Pach, R. Pollack, M. Sharir, Quasi-planar graphs have a linear number of edges. Combinatorica **17**, 1–9 (1997)
3. S. Avital, H. Hanani, Graphs. Gilyonot Lematematika **3**, 2–8 (1966) [in Hebrew]
4. A. Marcus, G. Tardos, Intersection reverse sequences and geometric applications. J. Comb. Theor. Ser. A **113**, 675–691 (2006)

5. J. Pach, Geometric graph theory, in *Surveys in Combinatorics, 1999*, ed. by J. Lamb, D. Preece. London Mathematical Society Lecture Notes, vol. 267 (Cambridge University Press, Cambridge, 1999), pp. 167–200
6. J. Pach, Geometric graph theory, in *Handbook of Discrete and Computational Geometry*, 2nd edition, ed. by J. Goodman, J. O'Rourke (Chapman & Hall/CRC, Boca Raton, FL, 2004) (Chapter 10)
7. J. Pach, R. Pinchasi, G. Tardos, G. Tóth, Geometric graphs with no self-intersecting path of length three. Eur. J. Combinator. **25**(6), 793–811 (2004)
8. J. Pach, R. Radoičić, G. Tóth, Relaxing planarity for topological graphs, in *Discrete and Computational Geometry*, ed. by J. Akiyama, M. Kano. Lecture Notes in Computer Science, vol. 2866 (Springer, Berlin, 2003), pp. 221–232
9. J. Pach, G. Tóth, Graph drawn with few crossings per edge. Combinatorica **17**, 427–439 (1997)
10. R. Pinchasi, R. Radoičić, On the number of edges in geometric graphs with no self-intersecting cycle of length 4, in *Towards a Theory of Geometric Graphs*. Contemporary Mathematics, vol. 342 (American Mathematical Society, Providence, RI, 2004), pp. 233–243
11. P. Valtr, Graph drawing with no k pairwise crossing edges, in *Graph Drawing* (Rome, 1997). Lecture Notes in Computer Science, vol. 1353 (Springer-Verlag, Berlin, 1997), pp. 205–218
12. G. Tardos, On 0-1 matrices and small excluded submatrices. J. Combinator. Theor. Ser. A **111**, 266–288 (2005)
13. G. Tardos, G. Tóth, Crossing stars in topological graphs. SIAM J. Discrete Math. **21**, 737–749 (2007)

A Better Bound for the Pair-Crossing Number

Géza Tóth

Abstract The crossing number $\text{CR}(G)$ of a graph G is the minimum possible number of edge crossings in a drawing of G, and the pair-crossing number $\text{PAIR-CR}(G)$ is the minimum possible number of crossing pairs of edges in a drawing of G. Clearly, pair-cr$(G) \leq$ cr(G). We show that for any graph G, cr$(G) = O(\text{pair-cr}(G)^{7/4} \log^{3/2}(\text{pair-cr}(G)))$.

1 Introduction

In a *drawing* of a graph G, vertices are represented by points and edges are represented by Jordan curves, in a plane, connecting the corresponding points. We assume that the edges do not pass through vertices, any two edges have finitely many common points, and each of them is either a common endpoint, or a proper crossing. We also assume that no three edges cross at the same point.

The *crossing number* $\text{CR}(G)$ is the minimum number of edge crossings (i.e., crossing points) over all drawings of G. The *pair-crossing number* $\text{PAIR-CR}(G)$ is the minimum number of crossing pairs of edges over all drawings of G. Clearly, for any graph G, we have

$$\text{PAIR-CR}(G) \leq \text{CR}(G).$$

It is still an exciting open question whether $\text{CR}(G) = \text{PAIR-CR}(G)$ holds for all graphs G.

Pach and Tóth [5] proved that $\text{CR}(G)$ cannot be arbitrarily large if $\text{PAIR-CR}(G)$ is bounded; namely, for any G, if $\text{PAIR-CR}(G) = k$, then $\text{CR}(G) \leq 2k^2$. Valtr [9] managed to improve this bound to $\text{CR}(G) \leq 2k^2/\log k$. Based on the ideas of Valtr, the present author [8] improved it to $\text{CR}(G) \leq 9k^2/\log^2 k$.

G. Tóth (✉)
Rényi Institute, Hungarian Academy of Sciences, Budapest, Hungary
e-mail: geza@renyi.hu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_30,

In this note, using a different approach, we obtain a further improvement.

Theorem. For any graph G, if PAIR-CR$(G) = k$, then CR$(G) = O(k^{7/4}\log^{3/2} k)$.

For the proof, we need some results about *string graphs*. These are introduced in Sect. 2. In Sect. 3, we give the short proof of the theorem. There are many other versions of the crossing number; for a survey, see [1, 6, 7].

2 String Graphs

A string graph is the intersection graph of continuous arcs in the plane. More precisely, vertices of the graph correspond to continuous curves (strings) in the plane such that two vertices are connected by an edge if and only if the corresponding strings intersect each other.

Suppose that $G(V,E)$ is a graph of n vertices. A *separator* in a graph G is a subset $S \subset V$ for which there is a partition $V = S \cup A \cup B$, $|A|, |B| \leq 2n/3$, and there is no edge between A and B. According to the Lipton–Tarjan separator theorem, [4], every planar graph has a separator of size $O(\sqrt{n})$. This result has been generalized in several directions, for graphs drawn on a surface of bounded genus, graphs with a forbidden minor, intersection graphs of balls in the d-dimensional space, intersection graphs of Jordan regions, intersection graphs of convex sets in the plane, and, finally, for string graphs [2, 3].

Theorem A ([3]). *There is a constant c such that for any string graph G with m edges, there is a separator of size at most $cm^{3/4}\sqrt{\log m}$.*

3 Proof of Theorem

Let c be the constant in Theorem A. In a drawing $\mathcal{D}$ of a graph G in the plane, call those edges that participate in a crossing *crossing edges*, and those that do not participate in a crossing *empty edges*.

Lemma. *Suppose that $\mathcal{D}$ is a drawing of a graph G in the plane with $l > 0$ crossing edges and $k > 0$ crossing pairs of edges. Then G can be redrawn such that (i) empty edges are drawn the same way as before, (ii) crossing edges are drawn in the neighborhood of the original crossing edges, and (iii) there are at most $6ck^{7/4}\log^{3/2} l$ edge crossings.*

Proof of Lemma. The proof is by induction on l. For $l = 1$, the statement is trivial. Suppose that the statement has been proved for all pairs (l', k'), where $l' < l$, and consider a drawing of G with k crossing pairs of edges, such that l edges participate in a crossing. Obviously, $\binom{l}{2} \geq k$, and $2k \geq l$; therefore, $2k \geq l > \sqrt{k}$.

Let V denote the vertex set of G and let E (resp., F) denote the set of empty (resp., crossing) edges of G. We define a string graph H as follows. The vertex set $\overline{F}$ of H corresponds to the crossing edges of G. Two vertices are connected by an edge if the corresponding edges cross each other. Note that the endpoints do not count; if two edges do not cross, the corresponding vertices are not connected even if the edges have a common endpoint. The graph H is a string graph; it can be represented by the crossing edges of G, as strings, with their endpoints removed. It has l vertices and k edges. By Theorem A, H has a separator of size $ck^{3/4}\sqrt{\log k}$. That is, the vertices can be decomposed into three sets, $\overline{F}_0$, $\overline{F}_1$, and $\overline{F}_2$, such that (1) $|\overline{F}_0| \leq ck^{3/4}\sqrt{\log k}$, (2) $|\overline{F}_1|, |\overline{F}_2| \leq 2l/3$, (3) there is no edge of H between $\overline{F}_1$ and $\overline{F}_2$.

This corresponds to a decomposition of the set of crossing edges F into three sets, F_0, F_1, and F_2, such that (1) $|F_0| \leq ck^{3/4}\sqrt{\log k}$, (2) $|F_1|, |F_2| \leq 2l/3$, (3) in drawing $\mathcal{D}$, edges in F_1 and in F_2 do not cross each other.

For $i = 0,1,2$, let $|F_i| = l_i$. Let $G_1 = G(V, E \cup F_1)$ and $G_2 = G(V, E \cup F_2)$. Then in the drawing $\mathcal{D}$ of the graph, G_i has l_i crossing edges. Denote by k_i the number of crossing pairs of edges of G_i in drawing $\mathcal{D}$. Then we have $k_1 + k_2 \leq k$, $l_1, l_2 \leq 2l/3$, and $l_1 + l_2 + l_0 = l$.

For $i = 1,2$, apply the induction hypothesis for G_i and drawing $\mathcal{D}$. We obtain a drawing $\mathcal{D}_i$ satisfying the conditions of the lemma: (1) Empty edges drawn the same way as before; (2) crossing edges are drawn in the neighborhood of the original crossing edges; and (3) there are at most $6ck_i^{7/4}\log^{3/2} l_i$ edge crossings.

Consider the following drawing $\mathcal{D}_3$ of G. (1) Empty edges are drawn the same way as in $\mathcal{D}$, $\mathcal{D}_1$, and $\mathcal{D}_2$; (2) for $i = 1,2$, edges in F_i are drawn as in $\mathcal{D}_i$; (3) edges in F_0 are drawn as in $\mathcal{D}$. Now count the number of edge crossings (crossing points) in the drawing $\mathcal{D}_3$. Edges in E are empty, edges in F_1 and in F_2 do not cross each other, and there are at most $2ck_i^{7/4}\log^{3/2} l_i$ crossings among edges in F_i. The only problem is that edges in F_0 might cross edges in $F_1 \cup F_2$ and each other several times, so we cannot give a reasonable upper bound for the number of crossings of this type. Color edges in F_1 and F_2 blue, and edges in F_0 red. For any piece p of an edge of G, let $\text{BLUE}(p)$ [resp., $\text{RED}(p)$] denote the number of crossings on p with *blue* (resp., *red*) edges of G. We will apply the following transformations.

REDUCECROSSINGS(e, f) Suppose that two crossing edges, e and f, cross twice, say, in X and Y. Let e' (resp., f') be the piece of e (resp., f) between X and Y. If $\text{BLUE}(e') < \text{BLUE}(f')$, or $\text{BLUE}(e') = \text{BLUE}(f')$ and $\text{RED}(e') \leq \text{RED}(f')$, then redraw f' along e' from X to Y. Otherwise, redraw e' along f' from X to Y. See Fig. 1.

Observe that REDUCECROSSINGS might create self-crossing edges, so we need another transformation.

REMOVESELFCROSSINGS(e) Suppose that an edge e crosses itself in X. Then X appears twice on e. Remove the part of e between the first and last appearances of X.

Start with drawing $\mathcal{D}_3$ of G, and apply REDUCECROSSINGS and REMOVE-SELFCROSSINGS recursively, as long as there are two crossing edges that cross at least twice, or there is a self-crossing edge.

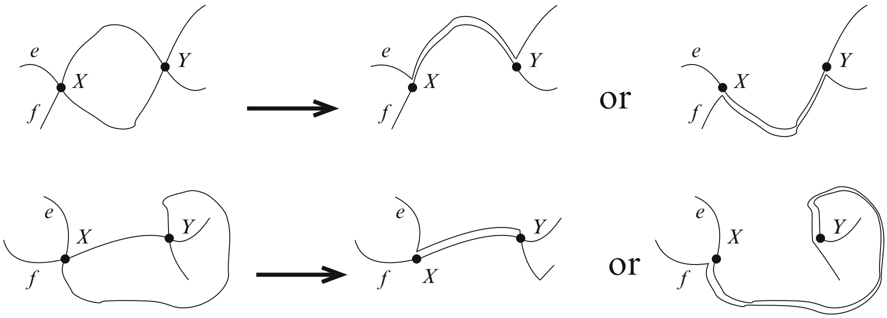


Fig. 1 REDUCECROSSINGS(e,f)

Let BB, (resp., BR, RR) denote the number of blue–blue (resp., blue–red, red–red) crossings in the current drawing of G. Observe that the triple (BB,BR,RR) lexicographically decreases with each of the transformations. Indeed,

- if e and f are both blue edges, then REDUCECROSSINGS(e,f) decreases BB,
- if e is blue and f is red, then either BB decreases, or if it stays the same, then BR decreases,
- if e and f are both red edges, then BB stays the same, and either BR decreases, or if it also stays the same, then RR decreases,
- if e is blue, then REMOVESELFCROSSINGS(e) decreases BB,
- and finally, if e is red, then BB does not change, BR does not increase, and RR decreases.

Therefore, after finitely many steps, we arrive at a drawing $\mathcal{D}_4$ of G, where any two edges cross at most once, and (BB,BR,RR) is lexicographically not larger than originally. That is, in the drawing $\mathcal{D}_4$, $BB \le 2ck_1^{7/4}\log l_1 + 2ck_2^{7/4}\log l_2$, and any two edges cross at most once; therefore, $BR + RR \le l_0 l$. So, for the total number of crossings, we have

$$\begin{aligned}
&6ck_1^{7/4}\log^{3/2} l_1 + 6ck_2^{7/4}\log^{3/2} l_2 + l_0 l\\
&\quad\le 6ck_1^{7/4}\sqrt{\log l}\log(2l/3) + 6ck_2^{7/4}\sqrt{\log l}\log(2l/3) + l_0 l\\
&\quad\le 6c(k_1^{7/4} + k_2^{7/4})\sqrt{\log l}(\log l + \log(2/3)) + l_0 l\\
&\quad\le 6ck^{7/4}\log^{3/2} l - 3ck^{7/4}\sqrt{\log l} + l_0 l\\
&\quad\le 6ck^{7/4}\log^{3/2} l - 3ck^{7/4}\sqrt{\log l} + clk^{3/4}\sqrt{\log k}\\
&\quad\le 6ck^{7/4}\log^{3/2} l - 3ck^{7/4}\sqrt{\log l} + 2ck^{7/4}\sqrt{\log k}\\
&\quad\le 6ck^{7/4}\log^{3/2} l - 3ck^{7/4}\sqrt{\log l} + 3ck^{7/4}\sqrt{\log l}\\
&\quad= 6ck^{7/4}\log^{3/2} l.
\end{aligned}$$

□

Now consider a graph G and let $\text{PAIR-CR}(G) = k$. Take a drawing of G with exactly k crossing pairs of edges. Let l be the total number of crossing edges. By the lemma, G can be redrawn with at most $6ck^{7/4}\log^{3/2} l$ crossings. Since $2k \geq l$, $\text{CR}(G) \leq 6ck^{7/4}\log^{3/2} l < 18ck^{7/4}\log^{3/2} k$. This concludes the proof of the theorem. □

Acknowledgements The author is very grateful to the anonymous referee for pointing out several typos and errors in the calculation, and for some other useful remarks.

The author was supported by OTKA Grants K-83767 and K-102029.

References

1. P. Brass, W. Moser, J. Pach, *Research Problems in Discrete Geometry* (Springer, New York, 2005)
2. J. Fox, J. Pach, Separator theorems and Turán-type results for planar intersection graphs. Adv. Math. **219**, 1070–1080 (2008)
3. J. Fox, J. Pach, A separator theorem for string graphs and its applications, in *WALCOM: Algorithms and Computation*. Lecture Notes in Computer Science, vol. 5431 (Springer, Berlin, 2009), pp. 1–14. Also in: Combinatorics, Probability and Computing, vol. 19 (2010), pp. 371–390
4. R.J. Lipton, R.E. Tarjan, A separator theorem for planar graphs. SIAM J. Appl. Math. **36**, 177–189 (1979)
5. J. Pach, G. Tóth, Which crossing number is it anyway? J. Combin. Theor. Ser. B **80**, 225–246 (2000)
6. J. Pach, G. Tóth, Thirteen problems on crossing numbers. Geombinatorics **9**, 194–207 (2000)
7. M. Pelsmajer, M. Schaefer, D. Štefankovič, Removing independently even crossings. SIAM J. Discrete Math. **24**, 379–393 (2010)
8. G. Tóth, Note on the pair-crossing number and the odd-crossing number. Discrete Comput. Geom. **39**, 791–799 (2008)
9. P. Valtr, On the pair-crossing number, in *Combinatorial and Computational Geometry*, Math. Sci. Res. Inst. Publ., vol. 52 (Cambridge University Press, Cambridge, 2005), pp. 569–575

Minors, Embeddability, and Extremal Problems for Hypergraphs

Uli Wagner

Abstract This is an expository chapter based on two talks given during the Special Semester on Discrete and Computational Geometry, organized by János Pach and Emo Welzl, at the École Polytechnique Fédérale in Lausanne, Switzerland, in the fall of 2010.

Our first purpose is to describe a circle of ideas regarding topological *extremal problems* for simplicial complexes (hypergraphs), and the corresponding *phase-transition* questions for *random complexes* (as introduced by Linial and Meshulam). In particular, we discuss some notions of *minors* pertaining to such questions. Our focus is on k-dimensional complexes that are *sparse*, i.e., such that the number of k-dimensional simplices is linear in the number of $(k-1)$-dimensional simplices.

Second, we discuss a notion of higher-dimensional *expansion* for simplicial complexes, due to Gromov, which is very useful in this context (the same notion of expansion has also arisen independently in the work of Linial, Meshulam, and Wallach on random complexes, and in the work of Newman and Rabinovich on higher-dimensional analogues of finite metrics).

1 Introduction

For graphs, there is a rich interplay between their combinatorial properties on the one hand and their topological properties on the other hand. A prime example is planarity. For instance, for (connected, finite) graphs embedded into the plane $\mathbf{R}^2$ (or the sphere $\mathbf{S}^2$), we have *Euler's relation* $f_0 - f_1 + f_2 = 2$ between the number f_0 of vertices, the number f_1 of edges, and the number f_2 of regions determined by

U. Wagner (✉)
Institut für Theoretische Informatik, ETH Zürich, CH-8092 Zürich Switzerland
e-mail: uli@inf.ethz.ch

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0_31,

the embedding. This implies the tight upper bound $f_1 \leq 3f_0 - 6$ for the number of edges of a *simple* planar graph with $f_0 \geq 3$ vertices.

Another fundamental result is the classical characterizations of planar graphs in terms of *forbidden minors*. Kuratowski [33] proved that a graph is planar iff it contains neither the complete graph K_5 nor the complete bipartite graph $K_{3,3}$ as a *topological minor* (i.e., if it does not contain a *subdivision* of either of these). Wagner [68] observed that as a rather immediate consequence, the same holds in terms of *deletion-and-contraction minors*; i.e., G is planar iff neither K_5 nor $K_{3,3}$ can be obtained from G by a finite sequence of edge or vertex deletions and edge contractions.

These theorems form the starting point of the general theory of graph minors, which by now comprises some of the deepest results and open problems in combinatorics, e.g., the Robertson–Seymour *graph minor theorem* [56] (see, e.g., [36] or [16, Chap. 12] for surveys) and *Hadwiger's conjecture* [23] (see also [5, 19]).

A related result of a very similar flavor that deserves to be mentioned here is the forbidden minor characterization of graphs that are *linklessly embeddable* into $\mathbf{R}^3$. Here, an embedding f of a graph G into $\mathbf{R}^3$ is called *linkless* if, for any two vertex-disjoint cycles in G, there exists a topological ball in $\mathbf{R}^3$ that contains the image of one cycle and is disjoint from the other one. Robertson et al. [55, 57] proved (confirming a conjecture of Sachs [58]) that a graph has such a linkless embedding iff it does not contain any minor that belongs to the *Petersen family*.[1]

Apart from structure theory, minors are also a valuable tool when dealing with extremal problems for *sparse graphs*, due to the following:

Theorem 1.4 (Mader [38]). *For every integer t, there are constants $C(t)$ and $C_{\mathrm{TOP}}(t)$ such that every (simple) graph on n vertices with at least $C(t) \cdot n$ edges [respectively, at least $C_{\mathrm{TOP}}(t) \cdot n$ edges] contains the complete graph K_t as a [topological] minor.*

Consequently, if $\mathcal{P}$ is a graph property for which there is a fixed *forbidden minor*[2] K, then graphs with that property have only a linear number of edges. This is the case, for instance, if $\mathcal{P}$ is nontrivial (i.e., if not all graphs have the property) and closed under taking minors. We stress, however, that we do not have to appeal to the graph minor theorem here: We do not need a full *characterization* in terms of forbidden minors but only the existence of some forbidden minor, a necessary condition that is usually much easier to prove.

For example, if one wishes to prove a linear upper bound for the number of edges in planar graphs without appealing to Euler's relation (we will see below why this is of interest with regard to higher-dimensional generalizations), then it is enough to

[1]By definition, the graphs in the Petersen family are the seven graphs that can be obtained from the complete graph K_6 by $Y\Delta$- or ΔY-*exchanges*. Here, a $Y\Delta$-exchange is the operation of deleting a vertex v of degree 3 and adding an edge between any two of the former neighbors, and a ΔY-exchange is the inverse operation.

[2]That is, no graph with property $\mathcal{P}$ contains K as a minor.

show two things: first, that K_5 (or some other fixed complete graph, if one is willing to accept a worse constant) is not planar, and, second, that planarity is preserved under taking minors (which is easy). The same type of argument also immediately implies a linear upper bound for the number of edges for graphs that are embeddable into a fixed surface, or linklessly embeddable into $\mathbf{R}^3$.

We remark that even the dependence on t of the constants in Mader's theorem is known very precisely: $C(t) \approx 0.319 \cdot t\sqrt{\log t}$ and $C_{\mathrm{TOP}}(t) = \Theta(t^2)$; see [65] and [6, 29], respectively. In the special case $t = 5$, Mader [39] also proved the sharp result that $3n - 5$ edges are already enough to ensure K_5 as a (topological) minor.

Higher Dimensions: Simplicial Complexes. It is rather natural to wonder if there are higher-dimensional analogues of graph minors. The first choice one faces in this context is what kind of objects to consider as higher-dimensional generalizations of graphs. Here, we focus on *simplicial complexes*. On the one hand, these can be defined and described completely combinatorially.[3] On the other hand, a simplicial complex also defines an *underlying topological space* for which it makes sense to consider *embeddability* into Euclidean spaces (generalizing planarity) and other topological properties.

We recall that a (finite, abstract) *simplicial complex* X is a finite set system that is closed under taking subsets; i.e., $F \subseteq G \in X$ implies $F \in X$. The sets in a simplicial complex X are called *faces* or *simplices* and are classified according to their *dimension* $\dim F := |F| - 1$. The dimension of X is defined as the maximum dimension of any face. The set of i-dimensional faces of X will be denoted by X_i, and the number of i-dimensional faces of X is denoted by $f_i(X) := |X_i|$, or simply by f_i. If $G \subseteq F \in X$, then we say that G is a face of F. The elements of a face F are called the *vertices* of F. Identifying singleton sets with their elements, we also call X_0 the set of vertices of X.

Thus, for example, a one-dimensional simplicial complex is just a simple graph, and f_0 and f_1 are the numbers of its vertices and edges, respectively.

Further basic notions regarding simplicial complexes, including *embeddings*, will be reviewed in Sect. 2.1.

In extremal and probabilistic combinatorics, it is somewhat more common to consider *hypergraphs*, i.e., arbitrary finite set systems, without the requirement of being closed under taking subsets. For our purposes, the two notions can be treated almost interchangeably, and the difference is rather one of viewpoint and terminology than an essential one.[4]

[3] As opposed to general *CW complexes*, say, for which one would also need to record the degrees of the attaching maps.

[4] Essentially, a *pure k-dimensional* simplicial complex (all maximal faces of dimension k) is the same thing as a $(k+1)$-uniform hypergraph (all sets in the set system, called *hyperedges*, are of cardinality $k+1$). Given an arbitrary hypergraph H, one can turn it into a simplicial complex by adding all subsets of the hyperedges in H to the hypergraph. Conversely, a simplicial complex is entirely determined by its maximal faces. One could say that, in a sense, the hypergraph viewpoint tends to focus on the interaction between the inclusion-maximal hyperedges and the vertices,

Minors. The second question is how to define a notion of minors for simplicial complexes. There are a number of possible definitions that are rather straightforward (and arguably too naive). Other, more refined, definitions have been proposed, notably by Nevo [49] and by Melikhov [46]. Which definition (if any) is "the right one" may depend on the context and on the aspects of the theory of graph minors that one wishes to generalize.

One very natural goal is to look for a higher-dimensional generalization of Kuratowski's planarity criterion, i.e., for a characterization of embeddability in terms of finitely many forbidden minors. The aforementioned papers by Nevo and in particular by Melikhov are steps in this direction. We remark, however, that recent computational hardness results regarding embeddability [42] give some indication that in full generality, the quest for a higher-dimensional generalization of Kuratowski's theorem may be too ambitious, at least for certain dimensions (e.g., for embeddings of two-dimensional complexes into $\mathbf{R}^4$).

Here, our main motivation is a related but somewhat different circle of problems, namely, topological *extremal problems* for *sparse* simplicial complexes and the corresponding *phase-transition* questions for random complexes (we will give more details shortly). In this context, we think of a simplicial complex as sparse if the number f_k of top-dimensional faces is linear in the number f_{k-1} of faces of codimension 1. In particular, this implies that $f_k \leq O(n^k)$, where $n = f_0$ is the number of vertices.[5]

In Sect. 5, we will discuss some of the difficulties one encounters when trying to use the existing notions of minors with regard to extremal questions, and we discuss a new notion of *homological minors* that was specifically designed as a tool for such applications.

Remark 2. The preprint by Melikhov [46] appeared after the original version of this chapter had been prepared and submitted, and it offers a notion of minors (for simplicial complexes as well as more general complexes) that subsumes, and is strictly more general than, that of Nevo. At present, we do not know whether the same difficulties that arise when trying to apply Nevo's notion of minors to extremal and phase-transition questions also arise for Melikhov's more general notion, although we suspect they do.

Extremal Questions. A prime example of the kind of topological extremal problem we have in mind is the following:

---

whereas the simplicial complex viewpoint stresses how hyperedges/faces "fit together" along faces of intermediate size.

[5]At the opposite end of the spectrum, one has *dense* simplicial complexes, for which $f_k \geq \Omega(n^{k+1})$. We remark that one could also consider more drastic definitions of sparsity, the most extreme of which would be to require that the complex be *locally bounded*, i.e., that the number of faces containing any given vertex should be bounded by an absolute constant, in which case one would have $f_i = O(n)$ for all i.

Conjecture 3. If a k-dimensional simplicial complex X embeds topologically into $\mathbf{R}^{2k}$, denoted $X \hookrightarrow \mathbf{R}^{2k}$, then $f_k(X) \leq C_k \cdot f_{k-1}(X)$, where C_k is a constant that depends only on k.

This is a natural generalization of the fact that the number f_1 of edges of a (simple) planar graph is at most linear in the number f_0 of vertices. The conjecture seems to be folklore and has appeared various times in the literature; see, e.g., [11, 27, 59]. Sometimes, the conclusion is weakened to $f_k = O(n^k)$, where $n := f_0$ is the number of vertices [we trivially have $f_{k-1} \leq \binom{n}{k}$ since a $(k-1)$-simplex has k vertices]. The best upper bound known to date for any fixed $k \geq 2$ is $O(n^{k+1-1/3^k})$; see Sect. 4.

As in the special case of graphs ($k = 1$), it is known that the general form of Conjecture 3 would have a number of interesting consequences in discrete geometry, including higher-dimensional analogues [11, 14] of the well-known crossing lemma [1, 34] and upper bounds for the number of triangulations of an n-point set in $\mathbf{R}^d$ [11, 15].

The problem is also closely related to the *upper bound theorem* and the *g-conjecture* for simplicial spheres, via a deep conjecture of Kalai and Sarkaria [27, 28, 59] that connects embeddability and algebraic shifting and would, in particular, imply Conjecture 3 in a very precise form. We discuss these implications and connections in Sect. 3.

We also remark that the case of k-dimensional complexes embeddable into $\mathbf{R}^{2k}$ is the critical one: On the one hand, every k-dimensional complex embeds into $\mathbf{R}^{2k+1}$. On the other hand, it is known that there are k-dimensional complexes that do not embed into $\mathbf{R}^{2k}$ (see Sect. 2.5).

Moreover, Conjecture 3 would imply asymptotically tight upper bounds for the face numbers of simplicial complexes embeddable into $\mathbf{R}^d$, for any ambient dimension d; see Sect. 3.

Random Complexes. Linial and Meshulam [35] introduced a higher-dimensional analogue of the Erdős–Rényi random graph model $G(n,p)$. By definition, the random k-dimensional complex $X^k(n,p)$ has n vertices, a *full* $(k-1)$-*skeleton* (i.e., every subset of the vertices that has size k or less forms a face of the complex), and every $(k+1)$-element set of vertices is taken as a k-face independently with probability p, which may be constant or, more generally, a function $p(n)$ depending on n.

This model has recently received a lot of attention, and the *threshold probabilities* for a number of topological properties of $X^k(n,p)$ have been determined [2, 3, 10, 31, 35, 47]; see Sect. 2.6.

It is known [37] that the sharp threshold probability for planarity of $G(n,p)$ is at $p = 1/n$. For the higher-dimensional embeddability problem, we have the following result.

Theorem 1.4 ([69]). *The (coarse) threshold for embeddability of $X^k(n,p)$ into $\mathbf{R}^{2k}$ is at $p = \Theta(1/n)$. More precisely, for every $k \geq 1$, there are constants $C_k > c_k > 0$ depending only on k such that*

$$\lim_{n\to\infty} \Pr[X^k(n,p) \hookrightarrow \mathbf{R}^{2k}] = \begin{cases} 1, & p \leq c_k/n, \\ 0, & p \geq C_k/n. \end{cases}$$

In particular, this shows that Conjecture 3 holds for almost all complexes.[6]

It seems very likely that, like planarity, the higher-dimensional embeddability property has a *sharp threshold*.

Conjecture 5. For every k, there exists a constant[7] C_k^* such that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr[X^k(n,p) \hookrightarrow \mathbf{R}^{2k}] = \begin{cases} 1, & p \leq (1-\varepsilon)C_k^*/n, \\ 0, & p \geq (1+\varepsilon)C_k^*/n. \end{cases}$$

Our current proof does not yield such a sharp threshold (even though many of the estimates could be improved, at the expense of making the proof more complicated).

On the other hand, our proof only uses a certain *quasi-randomness* or *expansion* property of $X^k(n,p)$, which will be discussed in Sect. 6, and thus applies to a broader class of complexes, which do not necessarily have a complete $(k-1)$-skeleton, including certain other models of random complexes. The relevant notion of expansion is due to Gromov [20]; the same notion arose independently in the work of Linial et al. [35, 47] on random complexes and in the work of Newman and Rabinovich [51] on higher-dimensional analogues of low-distortion embeddings of metric spaces.

The fact that $X^k(n,p) \hookrightarrow \mathbf{R}^{2k}$ asymptotically almost surely for $p \leq c_k/n$ follows fairly easily from a recent result of Aronshtam et al. [2] regarding the collapsibility of random complexes; see Sect. 2.6, Theorem 2.10. Thus, the main point is the proof of the nonembeddability part, which uses the notion of homological minors.

The Structure of This Chapter. The remainder of this chapter is structured as follows. In Sect. 2, we review some background concerning simplicial complexes, (co)homology, embeddings, obstructions, and random complexes.

In Sect. 3, we discuss some of the consequences that Conjecture 3, if true, would entail, such as a higher-dimensional analogue of the crossing lemma, and connections to other problems, such as the upper bound theorem and the g-conjecture for simplicial spheres.

In Sect. 4, we describe the general approach to extremal and threshold problems via forbidden minors. At its heart lies a conjectural higher-dimensional analogue of Mader's theorem. While we do not have a proof of this conjecture at the moment, as supporting evidence we have the corresponding threshold result for random

---

[6] For the complex $X^k(n,p)$, we have that $f_{k-1} = \binom{n}{k}$, and f_k is strongly concentrated around $p \cdot \binom{n}{k+1}$, so in order to obtain complexes with $f_k = C \cdot f_{k-1}$, the parametrization $p = (k+1)C/n$ is the right one.

[7] Computer experiments conducted by G. Pundak (personal communication) suggest that $C^*(2)$ is around 4.37.

complexes (Theorem 4.17). In Sect. 5, we survey several notions of minors for simplicial complexes, we give the precise definition of *homological minors*, and we show that, in a suitable technical sense, complexes with a nonembeddable homological minor are nonembeddable (Theorem 4.15). In Sect. 6, we discuss a higher-dimensional face expansion of simplicial complexes in the sense of Gromov and note that $X^k(n, C/n)$ has (a coarse version of) this property. Finally, in Sect. 7, we show that expanding complexes contain large complete minors (Proposition 7.32), which proves Theorems 1.4 and 4.17.

2 Preliminaries

We review some basic definitions and facts concerning simplicial complexes, homology, and obstructions in order to fix terminology and notation and to provide the necessary background for the definition of homological minors. For further background, see, e.g., [30, 40, 48].

2.1 Simplicial Complexes

Formally, there are two different ways of viewing a simplicial complex, either as an *abstract simplicial complex*, as defined in the introduction, or as a *geometric simplicial complex*. The latter means a finite collection X of closed geometric simplices in some Euclidean space $\mathbf{R}^m$ such that if $\sigma \in X$ and τ is a face of σ, then also $\tau \in X$, and such that any two simplices in X intersect in a common face (which may be empty). [Here, by definition, a geometric simplex σ is the convex hull of some affinely independent set of points, called the *vertices* of the simplex, and a face of σ is the convex hull of some subset (possibly empty) of its vertices.] Every geometric simplicial complex X defines an *underlying topological space* or *polyhedron* $|X| := \bigcup_{\sigma \in X} \sigma \subset \mathbf{R}^m$, namely, the union of all the geometric simplices in X, with the topology inherited as a subspace of the ambient Euclidean space $\mathbf{R}^m$.

Two simplicial complexes are isomorphic if there is a face-preserving bijection between their vertex sets. For any two isomorphic geometric simplicial complexes, there is an obvious homeomorphism between their underlying spaces that is linear on each face. There is a standard way of going back and forth between the abstract and geometric viewpoints (see, e.g., [40]), and an abstract simplicial complex can be viewed as a purely combinatorial description of a geometric simplicial complex up to isomorphism, by specifying which subsets of vertices form vertex sets of faces.

A *subcomplex* of X is a subset $Y \subset X$ that is itself a simplicial complex. The *i-skeleton* of X is the simplicial complex $X_{\leq i} = X_{-1} \cup X_0 \cup \ldots \cup X_i$ that consists of all faces of X of dimension at most i. A geometric simplicial complex X' is called a *subdivision* of a geometric simplicial complex X if $|X| = |X'|$ and every simplex of X' is contained in some simplex of X.

2.2 Maps and Embeddings

There are three natural notions of embeddings of a simplicial complex X into Euclidean space. In order of increasing generality, these are *linear embeddings*, *piecewise linear embeddings*, and arbitrary *topological embeddings*. We review the definitions below. For further background on embeddings of simplicial complexes, we refer the reader to the surveys [54,61]. For a discussion especially geared toward combinatorialists and computer scientists and with an emphasis on the algorithmic aspects of embeddability, see also [41].

Linear and PL Mappings of Simplicial Complexes. A *linear* mapping of a (geometric) simplicial complex X into $\mathbf{R}^d$ is a mapping $f\,|X| \to \mathbf{R}^d$ that is linear on each simplex. More explicitly, each point $x \in |X|$ is a convex combination $t_0v_0 + t_1v_1 + \cdots + t_sv_s$, where $\{v_0, v_1, \ldots, v_s\}$ is the vertex set of some simplex $\sigma \in X$ and $t_0, \ldots, t_s$ are nonnegative reals adding up to 1. Then we have $f(x) = t_0f(v_0) + t_1f(v_1) + \cdots + t_sf(v_s)$.

A mapping from X to $\mathbf{R}^d$ is *piecewise linear* (*PL*) if it is simplex-wise linear on some subdivision X' of X.

Embeddings. A (topological) *embedding* of a simplicial complex X into $\mathbf{R}^d$ is a map $f\colon |X| \to \mathbf{R}^d$ that is a homeomorphism of $|X|$ with $f(|X|)$. Since we only consider finite simplicial complexes, this is equivalent to requiring that f be injective. If such an embedding exists, we say that X embeds into $\mathbf{R}^d$, denoted $X \hookrightarrow \mathbf{R}^d$.

For a *PL embedding*, we require additionally that f be PL, and for a *linear embedding*, we are even more restrictive and insist that f be (simplex-wise) linear.

To illustrate the differences, consider the familiar example of embeddings of graphs in the plane. For a topological embedding, the image of each edge can be an arbitrary (curved) Jordan arc; for a PL embedding, it has to be a polygonal arc (made of finitely many straight segments); and for a linear embedding, it must be a single straight segment.

In the special case $\dim X = 1$ and $d = 2$ of graphs in the plane, all three notions happen to give the same class of embeddable complexes, namely, all planar graphs (by Fáry's theorem).

In higher dimensions, however, there are significant differences. For instance, Brehm and Sarkaria [7] showed that for every $k \geq 2$, and every d, $k+1 \leq d \leq 2k$, there is a k-dimensional simplicial complex X that PL embeds into $\mathbf{R}^d$ but does not admit a linear embedding.

It is also known that PL embeddability and topological embeddability do not always coincide. For example, there exists a four-dimensional simplicial complex (namely, the suspension of the Poincaré homology 3-sphere) that embeds topologically, but not PL, into $\mathbf{R}^5$ (Colin Rourke, private communication).

However, it is known that topological and PL embeddability do coincide for embeddings of k-complexes into $\mathbf{R}^d$ whenever the *codimension* $d-k$ is at least 3 [9], and also for $(k,d) = (2,3)$. The latter follows from results of Bing [4] and Papakyriakopoulos [52].

Here, we are mainly interested in embeddability in the topological sense (as opposed to linear embeddability, which is a much more geometric problem and has a very different “flavor”), and ”embeddability” will mean topological embeddability throughout, unless explicitly stated otherwise. For the most part, the subtle differences between PL and topological embeddability will play no role in our considerations. The sole exception is Proposition 3.13 below, where we need to assume PL embeddability in some cases.

Computational Aspects. One can also study the embeddability problem from a computational viewpoint. Specifically, in [41], the following computational decision problem $\mathrm{EMBED}_{k\to d}$ is considered for any fixed $d \geq k \geq 1$: Given a finite simplicial complex X of dimension at most k, does there exist a PL embedding of X into $\mathbf{R}^d$?

Known results imply that this problem is polynomial for $d \geq 2$ and in the case $d = 2k, k \geq 3$. However, the problem turns out to be algorithmically undecidable in certain cases, specifically, if $d \geq 5$ and $k \in \{d-1, d\}$, and to be at least NP-hard whenever k and d lie outside the *metastable range*, i.e., whenever $d \geq 4$ and $d \geq k \geq (2d-2)/3$. In particular, this is the case for $\mathrm{EMBED}_{2\to 4}$, the problem of embedding 2-complexes into $\mathbf{R}^4$. (It seems likely that the hardness proof carries over to topological embeddings, but the details have not been worked out.)

General Position and Perturbations. First, let’s consider a simplex-wise linear mapping f of a simplicial complex X into $\mathbf{R}^d$. We say that f is *in general position* if the images of any two vertices are distinct and the images of any $d+1$ or fewer vertices are affinely independent.

If f is a simplex-wise linear mapping in general position and if $\sigma, \tau \in X$ are disjoint simplices, then the intersection $f(\sigma) \cap f(\tau)$ is empty for $\dim \sigma + \dim \tau < d$ and has at most one point for $\dim \sigma + \dim \tau = d$.

A PL mapping of X into $\mathbf{R}^d$ is in general position if it corresponds to a linear mapping in general position on some subdivision of X. A PL embedding can always be brought into general position (by an arbitrarily small perturbation). A perturbation argument also immediately shows that any k-dimensional complex embeds into $\mathbf{R}2k+1$, even linearly. More explicitly, we can simply place the vertices in general position, for instance, on the moment curve, and interpolate linearly on the simplices; see, e.g., [40, Sect. 1.6]).

Moreover, the following observation will be important for what follows: If $f : X \to \mathbf{R}^d$ is a topological embedding, we can approximate f by a PL *almost-embedding*, i.e., a *PL* map $\tilde{f}$ such that the $\tilde{f}$-images of any two *vertex-disjoint* faces of X are disjoint. To see this, note that, by compactness, there exists some $\varepsilon > 0$ such that the f-images of any two vertex-disjoint faces have distance at least ε from each other. Take a sufficiently fine subdivision X' of X. For each vertex v of X', take $\tilde{f}(v)$ to be a point at distance at most $\varepsilon/2$ from $f(v)$ such that the resulting perturbed points $\tilde{f}(v)$ are in general position.[8] This defines a linear map on X' by interpolating

[8]For instance, choosing $\tilde{f}(v)$ uniformly at random from an ε^2-ball around $f(v)$ works with probability 1.

linearly on every simplex, and thus a PL map on X, and it is easy to see that it is a quasi-embedding.

2.3 Homology and Cohomology

We quickly review the definition of (reduced, simplicial) homology and cohomology of a finite simplicial complex X (see, e.g., [48] for a thorough introduction). For simplicity, we restrict ourselves to the case of coefficients in the field $\mathbf{F}_2$ with two elements (among other things, this allows us to ignore orientations and signs). For a thorough introduction to simplicial homology, we refer to any of the textbooks [30, 48, 53].

Let X be a finite simplicial complex. For integer i, denote by $C^i(X) = C^i(X;\mathbf{F}_2)$ the vector space $\mathbf{F}_2^{X_i}$ of functions from the set of i-faces of X to the field $\mathbf{F}_2$ [thus, $C^i(X) = 0$ unless $-1 \leq i \leq \dim X$]. The elements of this vector space are called *i-dimensional cochains* of X. Since we are working over $\mathbf{F}_2$, we can also think of an i-cochain as (the characteristic vector of) a subset of X_i.

Moreover, let $C_i(X) = C_i(X;\mathbf{F}_2)$ be the vector space over $\mathbf{F}_2$ generated by X_i. In other words, the elements of $C_i(X)$, called *i-chains*, are formal linear combinations of i-faces of X. Since the sets X_i are finite, $C_i(X)$ is again (noncanonically) isomorphic to $\mathbf{F}_2^{X_i}$, so it might seem like an exaggerated formalism to distinguish between $C_i(X)$ and $C^i(X)$, but it is sometimes convenient to maintain this distinction.

In fact, the space $C^i(X)$ is the dual vector space of $C_i(X)$, and we have a natural bilinear map $\langle , \rangle \colon C^i(X) \times C_i(X) \to \mathbf{F}_2$. If we identify both spaces with $\mathbf{F}_2^{X_i}$, this map corresponds to the standard inner product on $\mathbf{F}_2^{X_i}$.

We have a linear map $\partial = \partial_i : C_i(X) \to C_{i-1}(X)$, called the *boundary map*, given on basis elements $F \in X_i$ by $\partial F = \sum_{G \subseteq F, \dim G = i-1} G$. In other words, with respect to the standard bases of $C_i(X)$ and $C_{i-1}(X)$, the boundary map is given by the incidence (or inclusion) matrix between i-faces and $(i-1)$-faces.

Its dual map $\partial_i^* : C_{i-1}(X) \to C_i(X)$ is called the *coboundary map*. Again, with respect to the standard bases, this map is given by (the transpose of) the inclusion matrix. We often drop the indices and just write ∂ or ∂^*. Thus, if $S \subseteq X_i$ is a subset of i-faces and we view it as an i-chain, then the boundary ∂S corresponds to the set of all $(i-1)$-faces contained in an odd number of i-faces in S. Conversely, if we think of S as an i-dimensional cochain, then its coboundary consists of all $(i+1)$-faces that contain an odd number of faces in S.

The crucial property of the boundary map is that $\partial_{i-1} \circ \partial_i = 0$ (and consequently also $\partial_i^* \circ \partial_{i-1}^* = 0$), which is easily verified. Equivalently, we have the following relations between the kernels and images of these maps:

$$B_i = B_i(X;\mathbf{F}_2) := \operatorname{im} \partial_{i+1} \subseteq Z_i = Z_i(X;\mathbf{F}_2) := \ker \partial_i,$$

$$B^i = B^i(X;\mathbf{F}_2) := \operatorname{im} \partial_i^* \subseteq Z^i = Z^i(X;\mathbf{F}_2) := \ker \partial_{i+1}^*.$$

The elements of Z_i, B_i, Z^i, and B^i are called *i-cycles*, *i-boundaries*, *i-cocycles*, and *i-coboundaries*, respectively. By the above inclusion relations, we can form the quotient vector spaces $H_i(X) := \frac{Z_i(X)}{B_i(X)}$ and $H^i(X) := \frac{Z^i(X)}{B^i(X)}$, which are called the *i*-th *homology* and the *i*-th *cohomology group* of X, respectively.[9] Thus, every *i*-cycle $\zeta \in Z_i$ determines a *homology class* $[\zeta] = \zeta + B_i \in H_i$, and likewise every cocycle $\alpha \in Z^i$ determines a *cohomology class* $[\alpha] \in H^i$.

We also write $C_*(X) := \bigoplus_i C_i(X)$ and $C^*(X) := \bigoplus_i C^i(X)$ for the direct sums of the chain spaces and cochain spaces, respectively.

Example 6. A graph G is connected (for any two vertices, there is a path of edges from one to the other) iff $H_0(G) = 0$, which in turn is equivalent to $H^0(G) = 0$ [recall that we work with reduced (co)homology].

The first criterion $H_0(G) = 0$ is more or less directly equivalent to the definition of connectivity. Recall that over $\mathbf{F}_2$, an *i*-chain can be identified with a set of *i*-dimensional faces. For any pair of vertices $\{u, v\}$, if there is a path of edges between u and v, then the edges of this path form a 1-chain over $\mathbf{F}_2$ with boundary $\{u, v\}$. Conversely, any 1-chain with boundary $\{u, v\}$ is easily seen to contain a path.

The 0-cycles of G are exactly the even sets of vertices. Any such set can be obtained as a symmetric difference (sum over $\mathbf{F}_2$) of pairs. Thus, by the linearity of the boundary map, the above argument shows that any 0-cycle is the boundary of some 1-dimensional chain, i.e., $Z_0 = B_0$, and hence $H_0 = Z_0/B_0 = 0$.

On the other hand, the second criterion $H^0(G) = 0$ says that any function from the vertices of G to $\mathbf{F}_2$ that is locally constant along each edge (these are precisely the 0-cocycles) must be globally constant, i.e., a 0-coboundary, which is again equivalent to G being connected.

Simplicial (Co)Chain Maps and Induced Maps in (Co)Homology. If X and Y are simplicial complexes, then a *chain map* between them is a sequence of linear maps $\varphi_i : C_i(X) \to C_i(Y)$ with the additional property that these commute with the respective boundary maps in X and Y; i.e., $\varphi_i \circ \partial_i^X = \partial_i^Y \circ \varphi_{i-1}$. As before, we will often drop indices and write $\varphi(F)$ instead of $\varphi_i(F)$. Thus, the chain map represents every *i*-face F of X by a formal linear combination $\varphi(F)$ of *i*-faces in Y such that $\partial \varphi(F)$ equals the sum of the collections $\varphi(G)$, where G ranges over all $(i-1)$-faces in $\partial(F)$.

A chain map φ induces linear mappings $\varphi_* : H_i(X) \to H_i(Y)$ and $\varphi^* : H^i(Y) \to H^i(X)$ in (co)homology by defining $\varphi_*([\zeta]) := [\varphi_i(\zeta)]$ and $f^*([\alpha]) := [\varphi^i(\alpha)]$, where $\varphi^i : C^i(Y) \to C^i(X)$ is the transpose of φ_i.

A simplicial map f between complexes X and Y defines a chain map $f_\sharp$ that maps an *i*-face $F \in X_i$ to $f(F)$ if the latter is also of dimension i, and to zero otherwise. However, not all chain maps are of this form. One necessary condition is that a chain map induced by a simplicial map necessarily maps every vertex

[9]Thus, for the case $i = 0$, we work with what is sometimes called *reduced* (co)homology.

of X to a unique vertex of Y. Consequently, it maps the unique basis element $\emptyset$ of $C_{-1}(X)$ to the corresponding basis element $\emptyset$ of $C_{-1}(Y)$ (and not to zero). More generally, by a technique called *simplicial approximation*, one can show that an arbitrary continuous map $f : |X| \to |Y|$ between simplicial complexes induces homomorphisms (linear maps) $f_* : H_i(X) \to H_i(Y)$ and $f^* : H^i(Y) \to H^i(X)$ in (co)homology (note that the direction of the map is reversed in cohomology). Furthermore, homotopic maps induce identical homomorphisms in (co)homology.

On the level of chain maps, a *chain homotopy* P between two chain maps φ, ψ from X to Y is a family of linear maps (not chain maps) $P_i : C_i(X) \to C_{i+1}(Y)$ such that $\psi_i - \varphi_i = \partial^Y_{i+1} P_i + P_{i-1}\partial^X_i$ for all i. Chain-homotopic chain maps induce identical maps in (co)homology.

One fact we will need below as a black box is that the direct sum $H^*(X) = \bigoplus_i H^i(X)$ can be turned into a graded ring, called the *cohomology ring*. That is, one can define an associative, bilinear multiplication $H^i(X) \times H^j(X) \to H^{i+j}(X)$ called the *cup product*, and the homomorphisms induced by continuous maps also respect this product.

Moreover, in the discussion of deleted products below, we will need that everything said so far also applies to complexes more general than simplicial complexes. Specifically, we will need *polytopal* complexes, whose i-faces are i-dimensional convex polytopes that meet in common faces. The boundary matrix is still given by the incidences between i-faces and $(i-1)$-faces, and all other definitions carry over verbatim.

2.4 Deleted Products and Obstructions

Let X be a simplicial complex and $|X|$ its underlying topological space. The (twofold) *topological deleted product* of X is the space $|X|^2_{\mathrm{del}} := (|X| \times |X|) \setminus \{(x,x) : x \in |X|\}$, i.e., the twofold cartesian product with the "diagonal" removed.

The (twofold) *combinatorial deleted product* of X is the polytopal cell complex $X^2_{\mathrm{del}} := \{\sigma \times \tau : \sigma, \tau \in K, \sigma \cap \tau = \emptyset\}$. Thus, the cells of the deleted product are Cartesian products of vertex-disjoint simplices of X. The combinatorial deleted product is a subspace of the topological one. The latter *deformation retracts* onto the former; i.e., there is a map $F\colon |X|^2_{\mathrm{del}} \times [0,1] \to |X^2_{\mathrm{del}}|$ such that $F(\cdot, 0)$ is the identity, the image of $F(\cdot, 1)$ is contained in $|X^2_{\mathrm{del}}|$, and every $F(\cdot, t)$ fixes $|X^2_{\mathrm{del}}|$ pointwise. Consequently, the two spaces are homotopy equivalent.

The topological deleted product comes with an obvious action of the group $\mathbf{Z}_2$ that simply exchanges the order of coordinates, $(x,y) \mapsto (y,x)$. This is inherited by the combinatorial deleted product. For the latter, the action maps cells to cells, namely, $\sigma \times \tau \mapsto \tau \times \sigma$ (some care has to be taken regarding orientations, which we ignore here). The action is *free*; i.e., it does not have any fixed points. The homotopy equivalence between the topological and combinatorial deleted products can also be chosen to be equivariant. Henceforth, we will not distinguish between the two deleted products and simply write X^2_{del}.

An embedding $f\colon X \hookrightarrow \mathbf{R}^d$ induces a continuous map $\tilde{f}\colon X^2_{\mathrm{del}} \to S^{d-1}$ by setting $\tilde{f}(x,y) := \frac{f(x)-f(y)}{\|f(x)-f(y)\|}$. Moreover, this map is $\mathbf{Z}_2$-*equivariant*; i.e., $\tilde{f}(y,x) = -\tilde{f}(x,y)$ for all $(x,y) \in X^2_\Delta$. Thus, the existence of an equivariant map from X^2_{del} to S^{d-1} is a necessary condition for embeddability of X into $\mathbf{R}^d$.

In fact, one can conclude a bit more, using the fact that the topological and the combinatorial deleted product are equivariantly homotopy equivalent. Namely, if there exists no equivariant map from X^2_{del} to $\mathbf{S}^{d-1}$, then for any map $f\colon |X| \to \mathbf{R}^d$, there must be two *vertex-disjoint* faces of X whose images intersect.

A celebrated theorem by Haefliger and Weber [24, 70] asserts that for $\dim X \leq (2d-3)/3$ (called the *metastable range*), the existence of an equivariant map from X^2_{del} to $\mathbf{S}^{d-1}$ is also a sufficient condition for $X \hookrightarrow \mathbf{R}^d$ (outside the metastable range, this fails). We refer to [61] for a modern overview, proof sketch, and extensions.

Cohomological Obstructions. One can also formulate algebraic necessary conditions for the existence of equivariant maps. This yields the *van Kampen obstruction*. We first give a very quick (and rather abstract) definition (a special case of the general theory of bundles and classifying spaces) that allows us to derive the necessary facts quickly (for more details, see, e.g.[30, 45]).

It is a basic fact that there is a (cellular) $\mathbf{Z}_2$-equivariant map from X^2_{del} into the infinite-dimensional sphere $\mathbf{S}^\infty$. Moreover, this map is unique up to $\mathbf{Z}_2$-equivariant homotopy. In fact, both the map and the homotopy are easy to construct inductively on successively higher skeleta of X^2_{del}, using that $\mathbf{S}^\infty$ is contractible, i.e., that all its homotopy groups vanish. This $\mathbf{Z}_2$-map induces a unique map (up to homotopy) between the quotient spaces $X^2_{\mathrm{del}}/\mathbf{Z}_2 \to \mathbf{R}P^\infty$, and hence a unique homomorphism in cohomology $H^*(\mathbf{R}P^\infty;\mathbf{F}_2) \to H^*(X^2_{\mathrm{del}}/\mathbf{Z}_2;\mathbf{F}_2)$.

The $\mathbf{Z}_2$-cohomology ring of $\mathbf{R}P^\infty$ is known to be isomorphic to the polynomial ring $\mathbf{F}_2[x]$. In particular, in each dimension d, the element x^d is the unique nonzero element of $H^d(\mathbf{R}P^\infty;\mathbf{F}_2)$. The image of this element under the above homomorphism $H^d(\mathbf{R}P^\infty;\mathbf{F}_2) \to H^d(X^2_{\mathrm{del}}/\mathbf{Z}_2;\mathbf{F}_2)$ is called the van Kampen obstruction (modulo 2) for embeddability of X into $\mathbf{R}^d$ and denoted by $\mathfrak{o}^d_{\mathbf{F}_2}(X)$.

If X embeds into $\mathbf{R}^d$, then, as noted above, we get an equivariant map from X^2_{del} to $\mathbf{S}^{d-1}$. Composing this with the inclusion $\mathbf{S}^{d-1} \hookrightarrow \mathbf{S}^\infty$, we get a particular representative of the unique (up to homotopy) equivariant map from X^2_{del} into $\mathbf{S}^\infty$. Thus, on the level of cohomology, the induced map $H^*(\mathbf{R}P^\infty;\mathbf{F}_2) \to H^*(X^2_{\mathrm{del}}/\mathbf{Z}_2;\mathbf{F}_2)$ can be written as a composition

$$H^*(\mathbf{R}P^\infty;\mathbf{F}_2) \to H^*(\mathbf{R}P^{d-1};\mathbf{F}_2) \to H^*(X^2_{\mathrm{del}}/\mathbf{Z}_2;\mathbf{F}_2).$$

It is known that $H^*(\mathbf{R}P^{d-1};\mathbf{F}_2)$ is isomorphic to the quotient $\mathbf{F}_2[x]/(x^d)$ and that the map on the left is just the quotient map $\mathbf{F}_2[x] \to \mathbf{F}_2[x]/(x^d)$, so the image of x^d is zero. But then the image of x^d under the composed map is also zero; i.e., $\mathfrak{o}^d_{\mathbf{F}_2}(X) = 0 \in H^*(X^2_{\mathrm{del}}/\mathbf{Z}_2;\mathbf{F}_2)$.

A very simple but crucial observation that will be important later is that the existence and uniqueness up to homotopy of a cellular $\mathbf{Z}_2$-map can be mimicked on the level of chain maps, by using that all homology groups of $\mathbf{S}^\infty$ vanish.

Lemma 2.7. *Fix a scalar $a \in \mathbf{F}_2$ and a cellular decomposition of $\mathbf{S}^\infty$ that is compatible with the $\mathbf{Z}_2$-action. Then there is an $\mathbf{Z}_2$-equivariant chain map from X^2_{del} to $\mathbf{S}^\infty$ that maps $\emptyset \in C_{-1}(X^2_{del})$ to $a \cdot \emptyset \in C_{-1}(\mathbf{S}^\infty)$, and any two such chain maps are $\mathbf{Z}_2$-equivariantly chain homotopic.*

The proof is a very simple inductive argument, which we include below for the sake of completeness.

Since the van Kampen obstruction is the pullback of a cohomology class, it only depends on the chain homotopy class of the map used for the pullback. Thus, we obtain

Corollary 2.8. *Suppose there exists an equivariant chain map ψ from K^2_{del} to X^2_{del} such that $\psi(\emptyset) = 1 \cdot \emptyset$. Under this assumption, if $\mathfrak{o}^d_{\mathbf{F}_2}(K) \neq 0$ for some d, then also $\mathfrak{o}^d_{\mathbf{F}_2}(X) \neq 0$.*

Proof of Lemma 2.7. Note that, for each dimension $i \geq 0$, the $\mathbf{Z}_2$-action on X^2_{del} groups the i-faces into antipodal pairs. First, construct the chain map φ by induction on the dimension: Suppose it has been defined for all faces of dimension less than i. Pick an arbitrary representative i-face $F \times G$ from each antipodal pair of i-faces. By induction, the image $\varphi(\partial(F \times G))$ of the boundary is already defined, and since $\partial(F \times G)$ is a boundary, it is hence an $(i-1)$-cycle. Since the part of φ already defined is a chain map, it follows that $\varphi(\partial(F \times G))$ is an $(i-1)$-cycle as well. But $H_{i-1}(\mathbf{S}^\infty) = 0$, so there is an i-chain $\alpha \in C_i(\mathbf{S}^\infty)$ such that $\partial\alpha = \varphi(\partial(F \times G))$. Moreover, the part of φ defined already is equivariant; hence, $\varphi(\partial(G \times F)) = \nu(\varphi(\partial(F \times G))) = \nu(\alpha)$, where ν is the chain map induced by the antipodal $\mathbf{Z}_2$-action on $\mathbf{S}^\infty$. Thus, we can define $\varphi(F \times G) = \alpha$ and $\varphi(G \times F) = \nu(\alpha)$. Thus, inductively, we get the desired equivariant chain map.

Given two such chain maps φ and ψ that agree on the empty face, the desired equivariant chain homotopy between them is defined analogously by induction on the dimension. Here, we use the fact that when we need to define the image $P_i(F \times G)$ of an i-face $F \times G$, the summands of the chain $\zeta := \psi_i(F \times G) - \varphi_i(F \times G) - P_{i-1}(\partial(F \times G))$ are already defined, and this chain is a cycle since its boundary equals

$$\partial\zeta = \psi_{i-1}(\partial(F \times G)) - \varphi_{i-1}(\partial(F \times G)) - \underbrace{\partial P_{i-1}(\partial(F \times G))}_{=\psi_{i-1}(\partial(F\times G)) - \varphi_i(\partial(F\times G)) - P_{i-2}(\partial\partial(F\times G))} = 0,$$

where the last step follows by applying the chain homotopy condition to the chain $\partial(F \times G)$. Therefore, there exists an $(i+1)$-chain $\alpha \in C_{i+1}(\mathbf{S}^\infty$ with $\zeta = \partial\alpha$, and we can define $P_i(F \times G) := \alpha$. $\square$

A Concrete Description. The above definition of the van Kampen obstruction may seem rather abstract and unintuitive at first sight.

Here is a more concrete description (adapted from [41]) in the case $d = 2k$, which will be our main interest. Let X be a k-dimensional simplicial complex. Let

$$P := \{\{\sigma,\tau\} : \sigma,\tau \in X, \dim\sigma = \dim\tau = k, \sigma\cap\tau = \emptyset\}$$

be the set of unordered pairs of vertex-disjoint k-simplices in X. Likewise, let

$$Q := \{\{\sigma,\rho\} : \sigma,\rho \in X, \dim\sigma = k, \dim\rho = k-1, \sigma\cap\rho = \emptyset\}$$

be the set of pairs consisting of one k-simplex and a vertex-disjoint $(k-1)$-simplex.

We can identify P and Q with the sets of $2k$-faces and $(2k-1)$-faces of $X^2_{\mathrm{del}}/\mathbf{Z}_2$.

Now, we first define a particular vector $o_X \in \mathbf{F}_2^P \cong C^{2k}(X^2_{\mathrm{del}}/\mathbf{Z}_2)$. Fix an ordering of the vertices of X. We'll call a pair $\{\sigma,\tau\}$ of k-simplices *intertwined* if their vertices alternate according to the ordering of the vertices, i.e., $\sigma = [v_0, v_1, \ldots, v_k]$, $\tau = [w_0, w_1, \ldots, w_k]$, and $v_0 < w_0 < v_1 < w_1 < \cdots < v_k < w_k$ or $w_0 < v_0 < w_1 < v_1 < \cdots < w_k < v_k$. Then the vector o_X is defined componentwise by setting $(\mathfrak{o}_X)_{\{\sigma,\tau\}}$ to be 1 if σ and τ are intertwined, and zero otherwise.

Next, for each pair $\{\omega,\nu\} \in Q$, we define a vector $\varphi^{\omega,\nu} \in \mathbf{F}_2^P$ componentwise by setting $\varphi^{\omega,\nu}_{\sigma,\tau} = 1$ if $\sigma = \omega$ and $\nu \subset \tau$ or $\tau = \omega$ and $\nu \subset \sigma$, and $\varphi^{\omega,\nu}_{\sigma,\tau} = 0$ otherwise. These vectors are called *finger move vectors*.

We can now give an alternative definition of $\mathfrak{o}^{2k}_{\mathbf{F}_2}(X)$ (although it is not supposed to be obvious that the two definitions agree). Namely, $\mathfrak{o}^{2k}_{\mathbf{F}_2}(X)$ is the set of all vectors that can be obtained from o_X by adding a linear combination of finger move vectors. In other words, if we take the quotient of the vector space $\mathbf{F}_2^P$ by the subspace spanned by the finger move vectors, then $\mathfrak{o}^{2k}_{\mathbf{F}_2}(X)$ is the coset of o_X in that quotient space. In particular, we say that $\mathfrak{o}^{2k}_{\mathbf{F}_2}(X)$ vanishes or is zero if o_X is a linear combination of finger move vectors, which amounts to saying that an explicitly given system of linear equations over $\mathbf{F}_2$ is solvable.

Remark 9. Both the abstract and the concrete definition of the van Kampen obstruction treat the case of $\mathbf{F}_2$-coefficients. These definitions can be extended to integer coefficients, and it is known that for embeddings of k-complexes into $\mathbf{R}^{2k}$, $k \geq 2$, the integer-coefficient obstruction is strictly more powerful than the modulo 2 version; i.e., there are cases where the former can detect nonembeddability but the latter cannot (for the special case $k = 1$ of graphs, this does not happen, by the Hanani–Tutte theorem). In fact, for $k \neq 2$, vanishing of the integer-coefficient van Kampen obstruction actually yields a complete characterization of embeddability of a k-complex into $\mathbf{R}^{2k}$. We refer to [45] for more details about the van Kampen obstruction.

2.5 Small Nonembeddable Complexes

Every k-dimensional complex embeds into $\mathbf{R}^{2k+1}$. On the other hand, there are k-dimensional complexes that do not embed into $\mathbf{R}^{2k}$.

The basic examples are suitable *complete* and *complete multipartite* complexes, respectively. By definition, K_t^k, the *complete* k-dimensional simplicial complex on t vertices, has t vertices, and every subset of vertices of size at most $k+1$ forms a face. (In other words, K_t^k is the k-dimensional skeleton of the simplex on t vertices.) The *complete multipartite* complex $K_{t,\dots,t}^k$ has $t(k+1)$ vertices partitioned into $k+1$ classes $V_0,\dots,V_k$ of t vertices each, and the faces of $K_{t,\dots,t}^k$ are precisely the *rainbow* sets $F \subseteq V_0 \cup \ldots \cup V_k$, where F is rainbow if $|F \cap V_i| \leq 1$ for all i.

It is a classical result of van Kampen [67] and Flores [18] that the complexes K_{2k+3}^k and $K_{3,\dots,3}^k$ do not embed into $\mathbf{R}^{2k}$. This is proved by showing that the van Kampen obstructions of these complexes are nonzero.

These complexes are the direct generalizations of the nonplanar graphs K_5 and $K_{3,3}$. It is also known that the complexes K_{2k+3}^k and $K_{3,\dots,3}^k$ are *minimally nonembeddable*; i.e., if we remove from their underlying topological space an arbitrarily small open neighborhood of any point, then the resulting space becomes embeddable.

However, contrary to the situation of Kuratowski's theorem, these are not the only minimal nonembeddable complexes. For instance, any join $K_{2k_1+3}^{k_1} * K_{2k_2+3}^{k_2} * \cdots * K_{2k_s+3}^{k_s}$ that is of dimension k (this happens iff $k_1+k_2+\cdots+k_s+s-1=k$) is minimally nonembeddable into $\mathbf{R}^{2k}$ [21]. Furthermore, it is in fact known that for every $k \geq 2$, there is an infinite list of k-dimensional complexes that are not embeddable into $\mathbf{R}^{2k}$ and pairwise nonembeddable into each other [66, 71].

2.6 Random Complexes and Collapsibility

The Linial–Meshulam model $X^k(n,p)$ of k-dimensional random complexes described in the Introduction has been studied extensively, and the threshold probabilities for a number of topological properties have been determined quite precisely.

In particular, it is known that the sharp threshold for the vanishing of the $(k-1)$-cohomology $H^{k-1}(X^k(n,p);\mathbf{F}_2)$ with $\mathbf{F}_2$-coefficients is at $p=\frac{k\log n}{n}$; see [35, 47]. (The result extends to any fixed finite ring R of coefficients.)

Moreover, in the special case $k=2$, the threshold for the vanishing of the fundamental group $\pi_1(X^2(n,p)$ is roughly at $p=1/\sqrt{n}$; see [3].

In the top dimension, the coarse threshold for $H_k(X^k(n,p))=0$ (here, it actually does not matter which coefficients are used) is at $p=\Theta(1/n)$. That is, the following statements hold asymptotically almost surely (with probability tending to 1 as $n\to\infty$) (see [2, 10, 31]):

If $p = o(1)$, then $H_k(X^k(n,p)) = 0$, and if $p \geq c/n$, then $H_k(X^k(n,p)) \neq 0$, where $c = c(k)$ is a constant that depends only on k.

In fact, it is known that for $p = o(1/n)$, the complex $X^k(n,p)$ asymptotically almost surely *simplicially collapses* onto its $(k-1)$-skeleton [2, 10].

To define simplicial collapses, consider a k-dimensional complex X. A $(k-1)$-simplex $\sigma \in X$ is called *free* if there is a unique k-simplex τ containing it. We say that the subcomplex $X' := X \setminus \{\sigma, \tau\}$ arises from X by an *elementary collapse*[10] of the k-simplex τ through the $(k-1)$-simplex σ, and we say that X simplicially collapses onto a subcomplex Y if Y can be obtained from X by a sequence of elementary collapses.

One issue that arises in the case of the top-dimensional homology or collapsibility is that for any constant $c > 0$ and $p = c/n$, there is a constant probability that $X^k(n,p)$ will contain copies of K^k_{k+1} [the boundary of the $(k+1)$-simplex]. In fact, the number of such copies will be Poisson-distributed). If this happens, these simplex boundaries immediately create top-dimensional homology and prevent collapsibility. This poses a (minor) technical difficulty when trying to determine more precise thresholds.

To circumvent this, Aronshtam et al. [2] consider a slightly modified setting, where one conditions on the (constant probability) event that $X^k(n,p)$ does not contain any copies of K^k_{k+1}. Their results can also be rephrased as nonconditional statements. In particular, their proof yields the following result (although it is not explicitly stated in their paper).

Theorem 2.10 ([2]). *For every $k \geq 1$, there is a constant c_k such that $X^k(n, c_k)$ almost surely* simplicially collapses *to a subcomplex Y that consists of the $(k-1)$-skeleton together with some k-faces that form finitely many vertex-disjoint copies of K^k_{k+1}.*

Complexes as in the conclusion of this theorem are easily seen to PL embed into $\mathbf{R}^{2k}$. First, we linearly embed the disjoint simplex boundaries together with the $(k-1)$-skeleton (by general position). Then, by Lemma 2.11 below, we can greedily reverse the collapsing process one simplex at a time, until we get a PL embedding of the original complex.

Thus, the embeddability part of Theorem 1.4, i.e., the fact that $X^k(n,p)$ is almost surely embeddable for $p \leq c_k/n$, follows easily from Theorem 2.10.

Lemma 2.11. *Suppose X is a k-dimensional complex and that Y is a subcomplex that arises from X by an elementary collapse of a k-face τ through a free $(k-1)$-face σ. Then any PL embedding f of Y into $\mathbf{R}^{2k}$ can be extended to a PL embedding $\tilde{f}$ of X into $\mathbf{R}^{2k}$.*

The proof is not hard but a bit tedious and is omitted here. We just remark that the lemma is false, for instance, for embeddings of 2-complexes in $\mathbf{R}^3$. For instance,

[10] One can define free faces and elementary collapses more generally, but we will only need the special case discussed here.

the cone over K_5^1 collapses to a graph (the five edges connecting the five vertices of K_5^1 to the apex of the cone) but does not embed into $\mathbf{R}^3$ since, otherwise, the restriction of such an embedding to a small 2-sphere around (the image of) the apex would yield a planar embedding of K_5^1. Under the dimensional assumptions of the lemma, however, this kind of obstruction does not occur: If ρ is a proper face of τ present in Y and of dimension $\dim \rho = i$, say, then the given PL embedding of Y into $\mathbf{R}^{2k}$ restricts to an embedding of the $(k-i-1)$-dimensional complex $\mathrm{lk}(\rho, Y)$ into a small sphere $\mathbf{S}^{2k-i-1}$ linked with $f(\rho)$; by general position, this can be extended to an embedding of $\mathrm{lk}(\rho, X)$ into $\mathbf{S}^{2k-i-1}$. These embeddings can then be further extended to an embedding of X into $\mathbf{R}^{2k}$. We omit the details.

3 Some Consequences of Conjecture 3

In this section, we discuss some consequences of Conjecture 3 and some connections to related problems.

3.1 Higher-Dimensional Crossing Lemmas

The fact that the number of edges of a planar graph is at most linear has a number of important consequences, in particular the well-known *crossing lemma* [1,34], which asserts that for a simple graph with n vertices and $m \geq 4n$ edges, any plane drawing of the graph contains at least $\Omega(m^3/n^2)$ crossings.

One of the first applications of the crossing lemma was a singly exponential upper bound of B^n on the number of crossing-free geometric graphs on a given set P of n points in the plane [1], where B is some universal constant.[11]

Since then the Crossing Lemma has become a fundamental tool in discrete and computational geometry and many further applications have been found, most notably to Erdős distance problems [64] and to planar k-sets [12]. It also yields a very elegant proof of the Szemerédi–Trotter theorem on point-line incidences.

Conjecture 3 would have some analogous higher-dimensional consequences.

Theorem 3.12. *If* Conjecture 3 *is true, then the following statements hold as well:*

1. **High-dimensional crossing lemma**. *Let X be a k-dimensional simplicial complex with $f_k(X) > C \cdot n^k$. Then for any continuous map $f : X \to \mathbf{R}^{2k}$, there are at least $\Omega(f_k^{k+2}/f_{k-1}^{k+1})$ pairs of* vertex-disjoint *k-faces of X whose images under f intersect.*

---

[11] The original proof gave a very large base of the exponential ($B = 10^{13}$). Since then, this has been improved significantly, and it is known that $B \leq 344$ [60].

2. **Triangulations of point sets (Dey [11], Dey and Shah [15]).** *Let P be a set of n points in $\mathbf{R}^d$. Then the number of different triangulations of the point set P is at most $B^{n^{\lceil d/2 \rceil}}$, where $B = B(d)$ is a universal constant depending only on d. The same is true for the number of simplicial complexes* geometrically *embedded in the vertex set P.*

The proof of the first part is a straightforward application of the (by now) standard random sampling technique and is omitted here. Proving Conjecture 1 using the van Kampen obstruction modulo 2 would actually imply the same lower bound for the number of pairs of k-faces whose images cross an *odd* number of times. Using the weaker form $f_k = O(n^k)$ of Conjecture 1, a slightly weaker bound of $\Omega(f_k^{k+2}/n^{k(k+1)})$ for the number of crossing pairs was derived in [11]. See also [14] for further results of this kind.

It would be interesting to investigate whether some of the other applications of the crossing lemma can also be generalized to higher dimensions.

3.2 *The Upper Bound Theorem and Beyond*

Conjecture 3 would also imply tight bounds for the face number of complexes embeddable into other ambient dimensions.

Proposition 3.13. *Assume that Conjecture 3 holds for $k \le d/2$. Then, for every simplicial complex X that embeds*[12] *into $\mathbf{S}^d$, the face numbers of X satisfy $f_j(X) \le C_{j,d} \cdot f_{\lceil d/2 \rceil -1}(X)$ for $0 \le j \le \dim X$, where $C_{j,d}$ is a constant that depends only on j and d. In particular, the number of faces of X is at most $O(n^{\lceil d/2 \rceil})$, where $n = f_0$ is the number of vertices.*

This proposition is closely related to the upper bound theorem. McMullen's original theorem [43, 44] guarantees the conclusion of the Proposition 3.13 (with exact constants) for the special case that X not only embeds into $\mathbf{S}^d$ but forms the boundary complex of a convex $(d+1)$-polytope on n vertices. Stanley's generalization [62, 63] guarantees the same conclusion if X is a triangulation of $\mathbf{S}^d$, and Kalai's strong upper bound theorem [27] treats the case that X forms a subcomplex of the boundary of a $(d+1)$-polytope (but the polytope can contain more vertices than X).

Thus, Proposition 3.13 could be viewed as an asymptotic topological generalization of these results to complexes that are embeddable into the sphere (rather than triangulations of the sphere or simplicially embeddable into polytope boundaries).

The connections reach even farther, by a deep conjecture of Kalai and Sarkaria [27, 28, 59] that connects embeddability and algebraic shifting and would, in

[12] For technical reasons, in the cases $\dim X = d-1, d$, and $d > 3$, the proof only works if, instead of arbitrary topological embeddings, we consider *piecewise linear* embeddings.

particular, imply Conjecture 3 in a very precise form. See, e.g., [50, Sect. 6] for a more detailed discussion of this connection and further references to the literature.

The proof of Proposition 3.13 is a fairly simple induction over links of faces. Since we are not aware of a proof in the literature, we provide the details below. We remark that the conclusion of the proposition is known to be true unconditionally for d-dimensional complexes that embed *linearly* into $\mathbf{R}^d$; see [14].

Proof of Proposition 3.13. The proof is a fairly straightforward inductive argument based on considering links of faces (the basic idea, in the case $d = 3$, was already noted by Dey and Edelsbrunner [13]). The key observation is the following.

Observation 3.14. *If X embeds piecewise linearly into $\mathbf{S}^d$ and if F is an i-dimensional face of X, $-1 \leq i \leq \dim X$, then the link of F in X, denoted by $\mathrm{lk}(F,X)$, embeds piecewise linearly into $\mathbf{S}^{d-i-1}$.*

Suppose now that X embeds into $\mathbf{S}^d$. If $\dim X \leq d-2$ or $d \leq 3$n then it is known that there is also a piecewise linear (PL) embedding. If $\dim X = d-1, d$n and $d > 3$, we need to explicitly assume PL embeddability. In either case, choose and fix such a PL embedding.

To prove Proposition 3.13, we first note that if $\dim X < k$, then the assertion of the proposition is trivial, and if $\dim X = k$, then it follows immediately from Conjecture 3. Thus, we may assume that $\dim X > k$.

Suppose first that $\dim X$ is odd. Note that if $k = \lceil d/2 \rceil$, then $\lceil (d-1)/2 \rceil = k-1$. Consider a vertex $v \in X_0$. By the observation, we know that $\mathrm{lk}(v,X)$ embeds into $\mathbf{S}^{d-1}$. Thus, by induction on d, $f_{j-1}(\mathrm{lk}(v,X)) \leq C_{j-1,d-1} \cdot f_{k-1}(\mathrm{lk}(v,X))$ for $j = k+1, \ldots, \dim X - 1$. Thus,

$$(j+1)f_j(X) = \sum_{v \in X_0} f_{j-1}(\mathrm{lk}(v,X)) \leq C_{j-1,d-1} \sum_v f_{k-1}(\mathrm{lk}(v,X) = C_{j-1,d-1}(k+1) f_k(X);$$

i.e., $f_j(X) \leq C_{j,d} f_k(X)$, where $C_{j,d} = \frac{k+1}{j+1} C_{j-1,d-1}$.

Next, assume that $d = 2k$ is even. Consider an edge e of X. Again, by the observation, $\mathrm{lk}(e,X)$ embeds into $\mathbf{R}^{d-2} = \mathbf{R}^{2k-2}$. Thus, $f_{j-2}(\mathrm{lk}(e,X)) \leq C_{j-2,d-2} f_{k-2}(\mathrm{lk}(e,X))$, since $\lceil \frac{d-2}{2} \rceil - 1 = k-2$. Summing up over all edges, we obtain

$$\binom{j+1}{2} f_j(X) = \sum_{e \in X_1} f_{j-2}(\mathrm{lk}(e,X)) \leq C_{j-2,d-2} \sum_{e \in X_1} f_{k-2}(\mathrm{lk}(e,X)) = C_{j-2,d-2} \binom{k+1}{2} f_k(X).$$

Moreover, Conjecture 3 implies $f_k(X) \leq C_k f_{k-1}(X)$. Consequently, $f_j(X) \leq C_{j,d} f_{k-1}(X)$, where $C_{j,d} = C_k \cdot C_{j-2,d-2} \binom{k+1}{2} / \binom{j+1}{2}$. □

4 The Forbidden Minor Approach

The fact that the complete multipartite complex $K^k_{3,\dots,3}$ does not embed into $\mathbf{R}^{2k}$ immediately leads to a nontrivial upper bound $f_k = O(n^{k+1-1/3^k})$ for embeddable complexes, by the following Turán-type result.

THEOREM (ERDŐS [17]). *If a simplicial complex X on n vertices does not contain* $K^k_{t,\dots,t}$ *as a* subcomplex*, then* $f_k(X) = O(n^{k+1-1/t^k})$.

We remark that in higher dimensions, this extremal result is not known to be tight (see, e.g., the discussion in [22]). For $k \geq 2$, a probabilistic construction with alterations yields a slightly smaller exponent of $k+1-\frac{(k+1)t-k}{t^{k+1}-1}$ for the lower bound. Even so, forbidden subhypergraph arguments work well only for fairly dense complexes and are too weak to obtain tight bounds in our context. [To illustrate this, note that forbidding $K_{3,3}$ as a *subgraph* only gives an upper bound of $O(n^{5/3})$ for the number of edges of a planar graph, and an upper bound of $p \leq n^{-2/3}$ for the threshold of planarity of $G(n,p)$.]

A natural approach is to use *forbidden minors* instead. The challenge is to find a suitable notion of minors that, on the one hand, preserves sufficient topological information and, on the other hand, is sufficiently flexible so as to deal with very sparse complexes.

In Sect. 5, we briefly discuss some other notions of minors and the difficulties that arise when trying to carry out the above approach. The notion of *homological minors*, denoted by $\preccurlyeq_H$, is designed so as to circumvent these difficulties. As a first step, we show that homological minors satisfy the following version of the first requirement.

Theorem 4.15. *If* $K \preccurlyeq_H X$ *and* $\mathfrak{o}^{2k}_{\mathbf{F}_2}(K) \neq 0$*, then* $\mathfrak{o}^{2k}_{\mathbf{F}_2}(X) \neq 0$.

In other words: If there is a good reason for the nonembeddability of the minor K (nonvanishing of its van Kampen obstruction modulo 2), then X is nonembeddable for the same reason.[13]

Next, we propose the following generalization of Mader's theorem.[14]

Conjecture 16. For every $k,t \geq 1$, there is a constant $C = C(k,t)$ with the following property: If X is a simplicial complex with $f_k(X) \geq C \cdot f_{k-1}(X)$, then $K^k_t \preccurlyeq_H X$.

We have not yet been able to verify Conjecture 2 in full generality, but we can show that it holds for almost all complexes.

[13] There is an analogous result for the integer van Kampen obstruction, which, as remarked above, characterizes embeddability if $k \neq 2$.

[14] The case $t = 2k+3$ would be sufficient for the embeddability problem, but this special case seems to pose the same difficulties as the general one, so the restriction appears distracting rather than helpful. Moreover, there are other applications of homological minors that require larger complete minors. We also remark that up to a change in the constant, it does not matter whether one excludes complete or complete multipartite minors.

Theorem 4.17 ([69]). *For every k,t, there exists a constant $C = C(k,t)$ such that* $\Pr[K_t^k \preccurlyeq_H X^k(n,p)] \to 1$ *as $n \to \infty$ for $p \geq C/n$.*

The proof actually shows that with high probability, $X = X^k(n, C/n)$ contains a K_t^k-minor even after moderate *alterations*, i.e., if we remove εn^k many k-faces from X. This is of interest since random complexes $X^2(n,p)$ with small alterations show that an analogue of Conjecture 16 fails for topological minors of simplicial complexes (even if we allow a much larger probability, $p \approx 1/\sqrt{n}$); see Sect. 5.

Moreover, as remarked above, the proof only uses certain *quasi-randomness* or *expansion* properties of X and hence applies to a much broader class of complexes, which do not necessarily have a complete $(k-1)$-skeleton. We consider this as positive evidence that Conjecture 16, the notion of homological minors, and the overall approach proposed here are reasonable and promising.

Remark 18. In sparse (constant average degree) expanding graphs on n vertices, one can find not just constant-size complete minors, but much larger ones of size $\sqrt{n/\log n}$; see, e.g., [32]. The crucial fact that makes this possible is that such a graph has a very small diameter $O(\log n)$ and that this remains the case even if a moderate number of vertices are deleted. In higher dimensions, the picture is more complicated, and the higher-dimensional analogue of the diameter in an expanding complex need not be logarithmic; see [51]: There are examples of expanding 2-complexes X on n vertices that contain 1-chains that are triangle boundaries in the complete complex K_n^2 but are not the boundary of any 2-chain in X supported on fewer than $\Omega(n^{1/5})$ triangles.

5 Minors of Simplicial Complexes

We begin by discussing some possible ways of generalizing the notion of minors to simplicial complexes.

5.1 Subdivision and Topological Minors

From a topological point of view, maybe the most natural and straightforward thing to do is to generalize topological minors.

To do this, there are already several possibilities in higher dimensions. The most restrictive definition is the following. We say that a simplicial complex K is a *subdivision minor* of X, denoted $K \preccurlyeq_{\mathrm{SD}} X$, if X contains a subcomplex that is isomorphic to a subdivision of K.

We can weaken this in various ways. In a first step, we can require that for every face $F \in K$, X contain a subcomplex Y_F such that each Y_F is homeomorphic to F, i.e., to a ball of the appropriate dimension, and that the complexes Y_F fit together as required; i.e., $Y_F \cap Y_G = Y_{F \cap G}$. This is more general than the notion of subdivision minors, since for $\dim(F) = k \geq 5$, a complex Y_F may be a topological k-ball without being a PL-ball, i.e., without being isomorphic to a subdivision of F.

Next, we can further relax this and simply require that X contain a subcomplex Y that is homeomorphic to K, but without requiring that the homeomorphism $|K| \cong |Y|$ restrict to homeomorphisms on the faces of K.

Finally, we can simply require $|X|$ to contain a subspace (not necessarily a subcomplex) homeomorphic to $|K|$. All of these notions of minors obviously preserve embeddability: If K topologically embeds into X and X into $\mathbf{R}^d$, then K embeds into $\mathbf{R}^d$.

However, even the most permissive notion of topological minors fails when it comes to the existence of large complete minors in 2-complexes. Note that a two-dimensional complex X contains a subspace homeomorphic to $\mathbf{S}^2$ iff it contains a subcomplex homeomorphic to $\mathbf{S}^2$ (and since we are in dimension 2, that subcomplex will actually automatically be a PL 2-sphere, i.e., a subdivision of the boundary of a tetrahedron). Brown et al. [8] showed that there are two-dimensional complexes on n vertices with as many as $\Omega(n^{5/2})$ triangles that do not contain a subcomplex homeomorphic to the 2-sphere $\mathbf{S}^2$, i.e., that do not contain K_4^2 as a topological minor. Thus, the analogue of Conjecture 2 fails for these notions of minors.

A very simple alternative proof of this is due to Linial (personal communication) by a random construction with alterations: Consider the complex $X^2(n, c \cdot n^{-1/2})$; i.e., choose every possible triangle independently at random with probability $c \cdot n^{-1/2}$ (for some suitable constant $c > 0$). One can show that with high probability, the resulting complex contains few subcomplexes homeomorphic to $\mathbf{S}^2$, and these can be deleted by removing a triangle from each of them. The same construction can be used to exclude any surface of bounded genus.

As discussed above, for every $k \geq 2$, there are infinitely many k-dimensional complexes that are minimally nonembeddable into $\mathbf{R}^{2k}$ and do not embed into one another. This also shows that there is no hope for a higher-dimensional analogue of Kuratowski's theorem for topological minors.

As mentioned above, the embeddability problem $\mathrm{EMBED}_{k \to d}$ is also computationally hard in many cases, in particular in the case $\mathrm{EMBED}_{2 \to 4}$ of embeddability of 2-complexes into $\mathbf{R}^4$. We consider this as some discouraging evidence: If there is a higher-dimensional analogue of Kuratowski's theorem for 2-complexes embeddable into $\mathbf{R}^4$, then the corresponding notion of minors must be computationally hard (or maybe even undecidable). Of course, this does not rule out that such a notion of minors exists, but we expect that it might be hard to work with, in particular with regard to extremal questions.

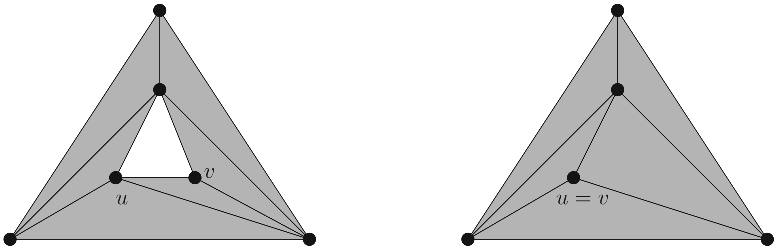


Fig. 1 A subdivided triangle with a hole contracts to a subdivided triangle

5.2 Deletions and Arbitrary Contractions

Another very natural attempt is to extend the commbinatorial definition of deletion-and-contraction minors and to allow arbitrary deletions and arbitrary contractions of faces of a simplicial complex.

Here, the most naive notion of a contraction would be to identify two vertices along a common edge. In the course of such a contraction of u with v, for every face $F = G \cup \{v\}$ in X with $u \notin G$, we remove F and replace it with the face $\mathbf{F}' = G \cup \{u\}$. If F' is already present in X, we just retain one copy (that is, we do not keep "multiple faces," as in edge contractions in simple graphs).

Let's say that K is a deletion-and-contraction minor of X, denoted by $K \preccurlyeq_{\mathrm{DC}} X$, if K can be obtained from X by a finite sequence of face deletions and edge contractions. One can show that this general notion of deletion-and-contraction minors satisfies an analogue of Conjecture 2 (essentially, the original proof of Mader [38] can be generalized to higher dimensions; we omit the details).

However, this contraction operation completely ignores the topology of the complex, and indeed, the first desired property for minors fails badly: We can have $K \preccurlyeq_{\mathrm{DC}} X$ and X embeds into $\mathbf{R}^d$ even though $\mathfrak{o}^d_{\mathbf{F}_2}(K) \neq 0$. For instance, consider the complex Y on the left-hand side of Fig. 1, which is homeomorphic to a triangle with a small open hole punched in its center. By contracting the edge uv, we obtain the complex Y' on the right, which is a subdivided triangle. Now consider the complete 2-complex K^2_7 on seven vertices. If we replace one of the triangles of K^2_7 with the complex Y, we obtain a complex X that is homeomorphic to K^2_7 with a small hole punched in one of the triangles. It is known (and easy to see) that X embeds into $\mathbf{R}^4$. However, if we contract the edge uv in X, we obtain a complex X' that is homeomorphic to K^2_7; hence, $\mathfrak{o}^4_{\mathbf{F}_2}(X') \neq 0$ and X' does not embed into $\mathbf{R}^4$. Thus, a notion of minors based on arbitrary contractions is unsuitable for embeddability questions.

One way around this problem is to leave the realm of simplicial complexes and to consider more general cell complexes in which cells are no longer uniquely determined by their boundary (higher-dimensional analogues of multigraphs). In this case, contracting the edge uv in the above example might look like Fig. 2.

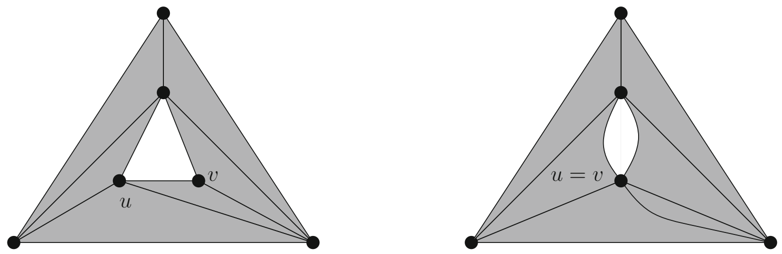


Fig. 2 Contractions in cell complexes may lead to parallel faces

So we would still retain information about nonembeddability. However, such cell complexes seem to be difficult to work with from the viewpoint of face numbers and extremal statements. For example, we cannot prove a linear upper bound for the number of edges in a planar multigraph in terms of the number of vertices. In order to prove any such statement, either we have to delete parallel edges, or we have to keep careful track of all multiplicities. We would need to do something similar for general cell complexes in higher dimensions, and it turns out to be quite tricky to define notions of "parallel faces" or "multiplicity" and to do the bookkeeping.

Moreover, this kind of contraction is still a special case of the more general homological minors defined below, which is why we prefer to work with the latter.

5.3 Admissible Contractions According to Nevo

Another way around the problem exemplified by Fig. 1 is to stay in the realm of simplicial complexes but to restrict the contractions that we allow. Nevo [49] recently introduced a notion of *admissible* contractions and a corresponding notion of minors. The idea is that a contraction as in the example above closes a "higher-dimensional hole" in the complex. In a certain sense, on the level of graphs this would correspond to identifying two vertices that are *not* joined by an edge, thus closing a zero-dimensional hole. To avoid this, Nevo considers *missing faces*: A k-face F of a complex X is *missing* if $F \notin X$, but all proper faces $G \subsetneq F$ belong to X. Identifying two vertices u ad v constitutes an *admissible contraction* in the sense of Nevo if u and v are not contained in a common missing face F.[15] Thus, in particular, if $\dim X \geq 1$, then u and v must form an edge (otherwise, the edge uv would be missing). If $\dim X \geq 2$, then for every common neighbor w of u and

[15]Nevo's original definition rules out only missing faces of dimension $\dim F \leq \dim X$, thus still allowing the possibility of u and v being contained in a common missing face of dimension $\dim X + 1$. As noted in the final version of [49], however, ruling out such missing faces as well leads to the same notion of minors.

v (i.e., a vertex such that both uw and vw are edges in X, also the triangle uvw must belong to the complex X), and so forth. A complex K is called a *deletion-and-admissible-contraction-minor*, denoted by $K \preccurlyeq_{\mathrm{DAC}} X$, if K can be obtained from X by a sequence of face deletions and admissible contractions. Nevo also showed that this notion of minors is well adapted to embeddability problems.

Theorem 5.19 (Nevo). *If $K \preccurlyeq_{\mathrm{Nevo}} X$ and if $\mathfrak{o}^d_{\mathbf{F}_2}(K) \neq 0$, then $\mathfrak{o}^d_{\mathbf{F}_2}(X) \neq 0$.*

However, it seems difficult to prove a variant of Conjecture 2 for this notion of minors. We do not have a proof that this is impossible, but the following considerations provide some evidence: Consider the random complex $X^2(n,p)$ on n vertices with a complete 1-skeleton, where each possible triangle is chosen independently with probability $p = C/n$ for some constant C. Then the probability that a given edge uv is incident to a certain number t of triangles equals $\binom{n-2}{t} p^t (1-p)^{n-2-t} \approx C^t \mathrm{e}^{-C}/t!$ for small t and $n \to \infty$. Thus, the triangle degree of a given edge is roughly Poisson-distributed. It follows that typically every edge is incident to many missing triangles, and it is not clear how to delete edges (and their incident triangles) in a controlled manner so as to obtain edges that are admissible to contract.

Remark 20. As mentioned in the Introduction, Melikhov [46] independently rediscovered the notion of admissible contractions and the corresponding notion of minors, which he calls *edge minors*. Moreover, he considers a more general notion of minors, which he calls *h-minors* (presumably, the "*h*" stands for "homotopy").

By definition, a simplicial complex K is an h-minor of a simplicial complex X if there exists a simplicial complex Y with $|Y| \subseteq |X|$ and a piecewise linear surjective map $f\colon |Y| \to |K|$ such that the preimage of every point in $|K|$ is *contractible*. Melikhov also considers a more restrictive version of minors, where every preimage is required to be *collapsible*. Let's call the latter, more restrictive version *sh*-minors (for "simple homotopy"). The following implications hold among these notions of minors: Every minor in the sense of Nevo is an *sh*-minor, and every *sh*-minor is an h-minor.

Due to time constraints in the preparation of the final version of the present chapter, we cannot give an adequate discussion of Melikhov's paper here. We just remark that Melikhov shows that for k-dimensional complexes, embeddability into $\mathbf{R}^d$ is preserved under taking h-minors if the *codimension* $d-k$ is at least 3, and embeddability is preserved unconditionally, for arbitrary codimensions, under taking *sh*-minors.

At present, we do not know whether an analogue of 4.17 might hold for h-minors.

5.4 Homological Minors

In order to circumvent the difficulties pointed out above, we propose the notion of homological minors.[16]

To motivate the definition, let's rephrase the definition of subdivision minors slightly differently. We have $K \preccurlyeq_{\mathrm{sd}} X$ iff, for every face $F \in K$, there is a subcomplex $Y_F \subseteq X$ that is isomorphic to a subdivision of F and such that $Y_F \cap Y_G = Y_{F \cap G}$ for all $F, G \in K$.

To define homological minors $K \preccurlyeq_H X$, we relax this in two ways: First, we drop the requirement that an i-face $F \in K$ be represented by a subcomplex $Y_F \subseteq X$ that is a combinatorial i-dimensional *ball* (isomorphic to a subdivided i-simplex). Instead, we allow more general i-dimensional chains. Second, we relax the conditions on how the Y_F may intersect.

Definition 5.21. *Let K and X be simplicial complexes. K is a* homological minor *of X, denoted $K \preccurlyeq_H X$, if*

0. there is a chain map *φ from K to X such that*
1. $\varphi(\emptyset) = 1 \cdot \emptyset$; equivalently, for every vertex v on K, its image $\varphi(v)$ is a set of vertices of odd *cardinality.*
2. φ preserves disjoint vertex supports*: If F and G are vertex-disjoint simplices, then every simplex in $\varphi(F)$ is vertex-disjoint from all simplices occurring in $\varphi(G)$.*

We remark that apart from the disjointness Condition 2, the definition boils down to linear algebra (which is the reason why chains are easier to handle than disks or other *homotopy-theoretic notions*): Condition 0 (being a chain map) amounts to a homogeneous system of linear equations, and Condition 1 to one inhomogeneous linear equation.

As an immediate consequence of the definition, we obtain the following.

Observation 5.22. *If φ is a chain map witnessing $K \preccurlyeq_H X$, then by Property 2, φ induces an equivariant chain map $\widetilde{\varphi}$ between the* deleted products *K^2_{del} and X^2_{del}, defined by "tensoring," i.e., $\widetilde{\varphi}(F \times G) = \sum_{A,B} A \times B$, where A ranges over the simplices in $\varphi(F)$ and B ranges over the simplices in $\varphi(G)$. Moreover, by Property 1, $\widetilde{\varphi}(\emptyset) = 1 \cdot \emptyset$; i.e., $\widetilde{\varphi}$ maps the empty face of K^2_{del} to that of X^2_{del}.*

---

[16] We mention that there is a line of research that studies generalizations of minors to matroids and related structures. For instance, we refer the reader to [19] for a survey of recent efforts to extend the graph minor theorem to matroids. Kaiser [26] further generalizes matroid minors to simplicial complexes, in the sense that independent complexes of matroids form a special class of simplicial complexes. These notions are somewhat related to the notion of homological minors, since the starting point for both is the linear algebra setting of cycles and cocycles (for a graph, simplicial complex, or matroid). However, matroid minors seem to focus only on top-dimensional cycles and to ignore the interactions between boundary maps of various dimensions, and they seem to have no direct bearing on embeddability and related questions.

Theorem 4.15 is a direct consequence of the preceding observation and Lemma 2.8.

Remark 23. In this chapter, we restrict ourselves to coefficients in the field $\mathbf{F}_2$, but the definitions carry over verbatim to arbitrary coefficient rings, with finite fields being most convenient to work with. In particular, the methods extend to Tverberg-type questions (where we are interested in p-fold covered image points instead of twofold ones), for which $\mathbf{F}_p$ is the appropriate choice of coefficients. One can also define a notion of minors based directly on the existence of chain maps between deleted products as in Lemma 2.8. However, in several ways, the original simplicial complex is technically easier to work with than the deleted product.

Remark 24. One can show that a minor in Nevo's sense is also a homological minor (the argument is implicit in [49]). However, homological minors are strictly more general. For instance, it is easy to show by induction on k that an arbitrary homology k-cycle X (considered a simplicial complex) contains K^k_{k+2} as a homological minor, but $K^k_{k+2} \preccurlyeq_{\text{Nevo}} X$ if and only if X is a piecewise linear k-sphere. A particularly simple example that shows that the two notions differ already for graphs (for which Nevo's minors are just usual graph minors) was pointed out by Nevo (personal communication): The "claw" $K^1_{1,3}$ contains K^1_3 (the boundary of a triangle) as a homological minor but not as a graph minor.

6 Higher-Dimensional Expansion

For graphs of arbitrary density, *edge expansion* can be defined as follows.

Definition 6.25. *Let $G = (V,E)$ be a graph, and let $\varepsilon > 0$. We say that G is ε-edge-expanding if, for every $S \subseteq V$,*

$$\frac{|E(S,V\setminus S)|}{|E|} \geq \varepsilon \cdot \frac{\min\{|S|,|V\setminus S|\}}{|V|}, \tag{1}$$

where $E(S,V\setminus S)$ is the set of edges across the cut $(S,V\setminus S)$, i.e., with one endpoint in S and the other one in $V\setminus S$.

[For graphs with bounded degrees, (1) is easily seen to be equivalent, up to a change in the constant ε, to the more usual condition that $|E(S,V\setminus S)| \geq \varepsilon|S|$ whenever $|S| \leq |V|/2$.] In order to generalize this to higher dimensions, we rephrase everything in terms of cochains.

Since we are working over $\mathbf{F}_2$, there is a one-to-one correspondence between subsets $S \subseteq V$ and 0-cochains (i.e., functions $\alpha\colon V \to \mathbf{F}_2$), by identifying S with its characteristic function $\mathbf{1}_S$, and the set $E(S,V\setminus S)$ of edges corresponds to the 1-cochain $\partial^*\alpha$. The constant 0-cochains $\mathbf{0}$ and $\mathbf{1}$ are precisely the coboundaries of the two possible (-1)-dimensional cochains, and they correspond to the trivial cuts with $S = \emptyset$ and $S = V$, respectively. Adding the constant zero-dimensional cochain

$\mathbf{1}$ to a 0-cochain $\alpha = \mathbf{1}_S$ is the same as exchanging the two sides S and $V \setminus S$ of the corresponding cut.

In general, let X be a simplicial complex. We equip the vector space $C^i(X) \cong \mathbf{F}_2^{X_i}$ with the *Hamming norm*; i.e., we define $|\alpha|$ to be the number of 1s that appear in the vector α, or equivalently, the number of i-faces in that are mapped to 1 by α. Then we normalize by the number of all i-faces and define $\|\alpha\| := \frac{|\alpha|}{f_i(X)}$.

In the case of 0-cochains, we defined expansion by bounding the norm $\|\partial^*\alpha\|$ of the coboundary from below in terms of $\min\{\|\alpha\|, \|\mathbf{1}+\alpha\|\}$. In general, the right measure is the *normalized distance* of α from the space B^i of coboundaries (the trivial kernel of ∂^*). That is, we define

$$\|[\alpha]\| := \min\{\|\alpha + \partial^*\beta\| : \beta \in C_{i-1}\}.$$

In other words, this is the quotient norm on the quotient space C^i/B^i induced by the normalized Hamming norm.

We remark that the definitions of $\|\alpha\|$ and $\|[\alpha]\|$ depend on the ambient complex [if also $\alpha \in C_i(Y)$ for some $Y \subseteq X$, then the norms may be different with respect to Y].

Now we are ready to define higher-dimensional expansion, which we refer to as *face expansion* or *cohomological expansion*. For our proof, we will also need a *coarse* version of it that only applies to large cochains.

Definition 6.26. *Let $\varepsilon > 0$. We say that a finite simplicial complex X has i-*dimensional face expansion ε *or that it is* ε-expanding in dimension i *if*

$$\|\partial^*\alpha\| \geq \varepsilon \cdot \|[\alpha]\| \tag{2}$$

holds for all $\alpha \in C^{i-1}(X)$. If we only require (2) for all $\alpha \in C^{i-1}(X)$ with $\|\alpha\| \geq \delta$ for some δ, then we say that X is coarsely (ε, δ)*-expanding in dimension i. We also call ε the* expansion factor *and δ the* coarseness.

Note that in the definition, we shifted the notation from i-chains to $(i-1)$-chains compared to the preceding discussion so that expansion in dimension i captures properties of the i-dimensional faces of X. For example, an ε-edge expanding graph is a simplicial complex that has one-dimensional face expansion ε.

The basic observation in this context (see [20, 35, 47]) is that the complete complex K_n^k is a face expander in all dimensions.

Proposition 6.27. *The complete complex K_n^k has i-dimensional face expansion* 1 *for all $i \in \{0, 1, \ldots, k\}$.*

Well-known Chernoff-type concentration bounds (see, e.g., [25]) easily imply the following.

Lemma 6.28 ([25]). *Let $Z = \sum_{i=1}^m Z_i$ be a random variable that is the sum of m independently and identically distributed random indicator variables with* $\Pr[Z_i = 1] = p = 1 - \Pr[Z_i = 0]$. *That is, Z has a binomial distribution with parameters n*

and p. Let $\lambda = np = \mathbf{E}Z$ be the expectation of Z. Then, for any $t \geq 0$,

$$\Pr[Z \leq \mathbf{E}Z - t] \leq \mathrm{e}^{-t^2/2\lambda}.$$

Lemma 6.29. *Let $X = X^k(n, C/n)$, where $C > 0$ is a sufficiently large constant. Then X is 1-expanding in every dimension $i \leq k-1$, and asymptotically almost surely X is coarsely expanding in dimension k with expansion factor $1/2$ and coarseness $\Omega(1/C)$.*

This is as an analogue of the fact that the random graph $G(n, C/n)$ has a giant component of size $\Omega(n)$ and that this component is edge-expanding.

Proof. The first assertion follows from Proposition 6.27.

For the second part, consider a $(k-1)$-chain α in $C_{k-1}(X)$ with $\|[\alpha]\| = a \geq \delta$, where δ is a parameter to be specified. We can also consider α a $(k-1)$-chain in the complete complex K_n^k containing X. In this bigger complete complex, the coboundary of α has size at least $a\binom{n}{k+1}$.

Since every k-simplex is chosen independently at random with probability $p = C/n$, the expected size of the coboundary of α in X equals $pa\binom{n}{k+1} \sim \frac{Ca}{k+1}\binom{n}{k}$. By Chernoff, the probability that the actual coboundary is less than half this large is at most $\mathrm{e}^{-\frac{Ca}{8(k+1)}\binom{n}{k}}$. On the other hand, the total number of cochains α is at most $2^{\binom{n}{k}}$. Thus, if δ is at least $8(k+1)\ln(2)/C$, a union bound shows that asymptotically almost surely, every α with $\|[\alpha]\| \geq \delta$ satisfies $|\partial^*\alpha| \geq \frac{1}{2}ap\binom{n}{k+1}$.

Moreover, the total number of k-simplices in X is tightly concentrated around $p\binom{n}{k+1}$; i.e., asymptotically almost surely $f_k(X) = (1+o(1))p\binom{n}{k+1}$, and hence $\|\partial^*\alpha\| \geq 1/2(1+o(1))\|[\alpha]\|$ holds whenever $\|[\alpha]\| \geq \delta$. □

We remark that the expansion factor of $1/2$ is chosen completely arbitrary. We could choose any other fixed constant $\varepsilon < 1$, at the expense of making the constant of proportionality between the coarseness δ and $1/C$ worse. The same argument also shows that $X^k(n, \frac{C\log n}{n})$ is expanding in dimension k if $C = C(k)$ is a sufficiently large constant.

7 Partitions and Colorful Homological Minors and Cominors

Let K and X be simplicial complexes of dimension k. Consider a graded linear map $\varphi : C_*(K) \to C_*(X)$, i.e., a family of linear maps $\varphi_i : C_i(K) \to C_i(X)$. We can identify each φ_i with a matrix of dimension $f_i(X) \times f_i(K)$ over $\mathbf{F}_2$, and by writing the entries of the matrices in a row (in some specified order), we can view φ as a vector over $\mathbf{F}^2$ of length $\sum_i f_i(X) f_i(K)$. The entries of this vector are indexed by pairs of faces $(A, F) \in K \times X$ of equal dimension $\dim A = \dim F$, and the entry of φ in position (A, F) equals the coefficient with which F appears in the chain $\varphi(A)$.

Condition 0 in the definition of homological minors, the property of being a chain map, simply amounts to a number of homogeneous linear conditions on the entries of φ, with the coefficients of the linear equations given by the boundary matrices of K and X.

Condition 1, $\varphi(\emptyset) = \emptyset$, is also a linear condition, albeit an inhomogeneous one: It simply says that a fixed entry of φ (w.l.o.g. the first entry) equals 1. Only Condition 2, disjointness of vertex supports, is not linear. However, we can enforce it as follows.

7.1 Partitions and Colorful Chain Maps

Suppose that K has t vertices, w.l.o.g. $K_0 = \{1,\dots,t\} =: [t]$. Fix a partition of the vertices of X into t parts or *color classes*, in other words, a map $\mathcal{P} : X_0 \to [t]$. We restrict our attention to *colorful* simplices of X, i.e., faces $F \in X$ that contain at most one vertex of each color.

For a given $A \in K$, let's say that a simplex $F \in X$ is A*-colored* if F contains precisely one vertex of each color $i \in A$ (such a simplex is necessarily of the same dimension as A), and let $X[A]$ denote the set of A-colored simplices in X (strictly speaking, $X[A]$ depends on the partition $\mathcal{P}$, but we suppress this from the notation). Let's say that a graded map $\varphi : C_*(K) \to C_*(X)$ is *color-faithful* if, for every $A \in K$, the chain $\varphi(A)$ is supported on the A-colored simplices $X[A]$; i.e., every simplex $F \in X$ that appears with a nonzero coefficient in the chain $\varphi(A)$ is A-colored. Such a φ trivially preserves vertex disjointness, since for vertex-disjoint $A, B \in K$, any $F \in X[A]$ and $G \in X[B]$ are vertex-disjoint as well.

From now on, we only consider color-faithful graded linear maps $\varphi : C_*(K) \to C_*(X)$. As before, we can identify such a map with a vector over $\mathbf{F}_2$, whose entries are indexed by pairs (A,F), where $A \in K$ and F is now required to be A-colored. Thus, the length of φ equals $\sum_{A\in K} |X[A]|$. Condition 0 of being a chain map means that φ is a solution to a homogeneous system of linear equations,

$$M\varphi = 0. \tag{3}$$

Here, M is called the *enhanced boundary matrix of* X and defined as follows. The columns of M are indexed by pairs $(A,F) \in K \times X$ with $F \in X[A]$ as above, and the rows of M are indexed by pairs (G,i), where G is a colorful face of X *of dimension at most* $(k-1)$, G is B-colored for some $B \in K$, and $i \in [t] \setminus B$ is a color that does not appear in G. The entry of M in row (G,i) and column (A,F) equals 1 if either $A = B \cup \{i\}$ and $G \subseteq F$ (i.e., if G is a facet of F and i is the color of the unique vertex of F missing from G) or if $G = F$. Otherwise, the entry of M at that position is zero. See Fig. 3, for a schematic illustration of M in block form.

Remark 30. For a fixed K and a *random partition* of the vertices of X, the existence of a nonzero solution to (3) is almost trivial if $f_k(X) \geq C \cdot f_{k-1}(X)$ for a sufficiently

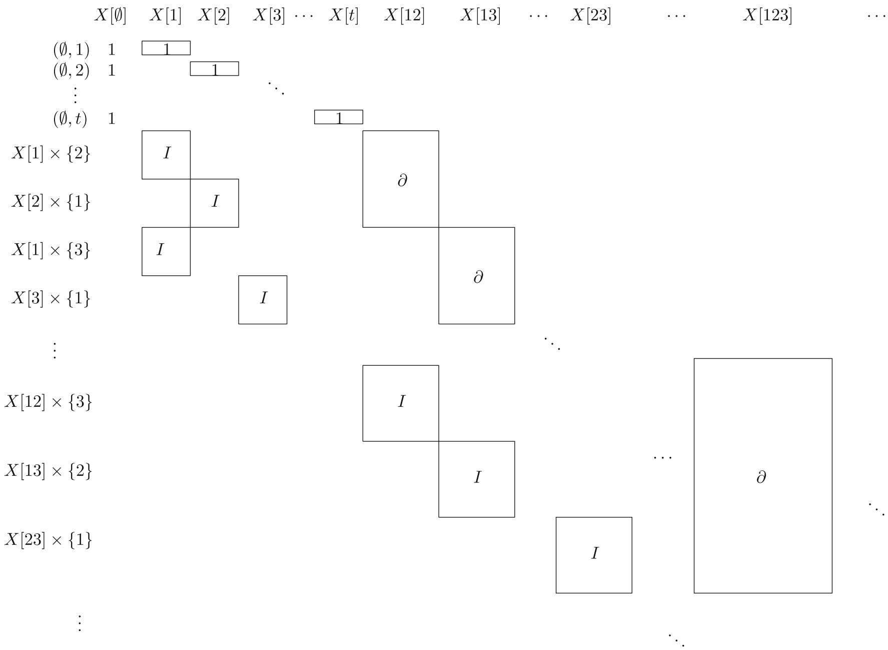


Fig. 3 The (*upper left corner*) of the enhanced boundary matrix for the case $K = K_t^k$. Here, we have grouped the row indices (G,i) according to which $X[B]$ the face G belongs to, and the column indices (A,F) according to A. Moreover, each I stands for the appropriate identity matrix on $X[A]$, and each ∂ stands for the appropriate portion of the boundary matrix of X (restricted to the entries indexed by $X[B] \times X[A]$)

large constant C, by counting variables and constraints: For the rows of M, we only consider faces of X of dimension $i < k$, and each such face appears only $t-i-1$ times. Thus, the total number of rows (linear equations) of M is at most $\sum_{i=-1}^{k} -1(t-i-1)f_i(X) \leq t2^k f_{k-1}(X)$. On the other hand, the number of columns (variables) of M is at least the number of *colorful* k-simplices of X. If we pick a random partition, then the expected number of colorful k-simplices is $\frac{1}{(t-1)(t-2)\cdots(t-k)} f_k(X) \geq t^{-k} C f_{k-1}(X)$, and for sufficiently large C, this is larger than the number of rows, so M has a nontrivial kernel.

However, we are looking for a solution φ of (3) that also satisfies one additional *inhomogeneous condition* $\varphi(\emptyset) = \emptyset$, i.e., for a vector $\varphi \in \ker M$ whose entry in the first position equals 1. Because of this inhomogeneity, the simple dimension-counting argument breaks down (two affine subspaces need not intersect, even when the sum of their dimensions exceeds the dimension of the ambient space).

7.2 Cominors

When does a solution $\varphi \in \ker M$ with first coordinate equal to 1 exist? It does *not* exist if and only if the row vector $(1,0,\ldots,0)$ lies in the row space of M i.e., if there is a row vector α such that $\alpha M = (1,0,\ldots,0)$. We call any such row vector α a *cominor*.

Let's take a closer look at what it means. The entries of α are indexed by the same pairs (G,i) as the rows of M; i.e., G is a colorful face of X of dimension $\dim(G) < k$, and $i \in [t]$ is a color that does not appear in G. The conditions for α to be a cominor are as follows: First of all, $\sum_{i=1}^{t} \alpha(\emptyset, i) = 1$ (all calculations are modulo 2). Second, for every nonempty colorful face F of X,

$$\sum_{i\in[t]\setminus\mathcal{P}} \alpha(F,i) \;+ \sum_{G\subset F, |G|=|F|-1} \alpha(G,\mathcal{P}(F\setminus G)) = 0$$

if $\dim F < k$ and

$$\sum_{G\subset F, |G|=|F|-1} \alpha(G,\mathcal{P}(F\setminus G) = 0$$

if $\dim F = k$.

Equivalently, we can view α as a collection $\{\alpha_A\}$ of $\dim(A)$-dimensional cochains on the subcomplexes[17] $X[A]$, one for each simplex $A \in K$, by setting $\alpha_A(F) := \alpha(F,i)$, where F is face of X that is B-colored for the unique facet $B \subset A$ with $\{i\} = A \setminus B$; i.e., i is the unique color of A that does not appear in F.

Thus, in the case $K = K_t^k$, for each $i \in [t]$, we have a (-1)-dimensional cochain α_i on the zero-dimensional complexes $X[i]$. For each pair $\{i,j\} \in \binom{[t]}{2}$, we have a 0-cochain on $X[\{i,j\}]$, etc. The condition of being a cominor translates to the requirement that $\sum_{i=1}^{t} \alpha_{\{i\}}(\emptyset) = 1$, and for $A \in K_t^k$ and $F \in X[A]$,

$$\partial^* \alpha_A(F) = \sum_{i\in[t]\setminus A} \alpha_{A\cup i}(F) \tag{4}$$

if $0 \le \dim A < k$ and

$$\partial^* \alpha_A = 0 \tag{5}$$

if $\dim A = k$. Thus, in the latter case, α_A is a *cocycle* on $X[A]$ (note that α_A need not be a cocycle when considered a chain in X).

In order to show the existence of a K_t^k minor, we will assume there is a cominor and derive a contradiction. For this, we use face expansion.

---

[17] Here, we abuse notation and think of $X[A]$ as the subcomplex consisting of all A-colored simplices and their faces.

7.3 Minors in Expanding Complexes

We are now ready to prove Theorem 4.17. The first step is the following lemma.

Lemma 7.31. *Let X be a k-dimensional simplicial complex with a given t-partition of the vertices, and assume that α is a K_t^k-cominor with respect to that partition. Then we may modify α, while preserving the cominor conditions, and achieve that for every $A \in K_t^k$, the cochain α_A is* cohomologically minimal, *i.e.* $\|\alpha_A\| = \|[\alpha_A]\|$, *in the subcomplex $X[A]$.*

Proof. We prove this top-down. Let $i \leq k$ and $C \in \binom{[t]}{i+1}$. Then the component α_C of the cominor is an $(i-1)$-dimensional cochain on $X[C]$. Let β be an arbitrary $(i-1)$-dimensional coboundary on $X[C]$. It suffices to show that we can modify α to obtain a new cominor α' such that $\alpha'_C = \alpha_C + \beta$ and $\alpha'_A = \alpha_A$ for all $A \subseteq [t]$ with $|A| \geq i+1$ and $A \neq C$. (If we can prove this, then we can make all α_A cohomologically minimal in a top-down fashion, since our next modification never interferes with any of the previous ones in the same or higher dimensions.)

Let γ be an $(i-2)$-dimensional cochain on $X[C]$ with $\partial^*\gamma_C = \beta_C$. For every "facet" $B \in \binom{C}{i}$ of C, define $\alpha'_B := \alpha_B + \gamma|_{X[B]}$ (each $X[B]$ is a subcomplex of $X[C]$, so the restriction of γ is defined). Moreover, define $\alpha'_C := \alpha_C + \beta$, and set $\alpha'_A := \alpha_A$ for every $A \subseteq [t]$ that is not equal to C or a facet of C. We claim that the resulting α' is still a cominor.

Consider a subset $A \subseteq [t]$. If $A = C$, then $\partial^*\alpha'_C = \partial^*(\alpha_C + \beta) = \partial^*\alpha_C$, and $\alpha'_S = \alpha_S$ for all proper supersets $S \supset C$. Therefore, whichever of the two cominor relations (4) and (5) applies, it is preserved.

Next, suppose that A is a facet of C and let $F \in X[A]$. We have $\partial^*\alpha'_A(F) = \partial^*\alpha_A(F) + \beta(F)$. On the other hand, there is a unique $i \in [t] \setminus A$ such that $A \cup i = C$. For this i, we have $\alpha'_{A\cup i}(F) = \alpha'_C(F) = \alpha_C(F) + \beta(F)$. For all other $j \in [t] \setminus A$, $A \cup j$ is neither equal to C nor to a facet of C; hence, $\alpha'_{A\cup j}(F) = \alpha_{A\cup j}(F)$ for all these other j. Summing up, we see that relation (4) is preserved since we simply add $\beta(F)$ once on either side of the equation.

If $A \subseteq C$ and $|A| = i-1$, then there are precisely two superfaces $B \subset A$ of size $|B| = i$ that are facets of C. On each of these two facets, α_B is modified by adding γ. On all other B, α_B is unchanged. Thus, in the cominor relation (4) for A, we add $\beta(F)$ exactly twice to the right-hand side and make no changes on the left-hand side, so the relation is preserved.

For all other A, neither α_A nor the cochains $\alpha_{A\cup i}$ are affected. This shows that the modified α' is a cominor, which completes the proof of the lemma. □

Proposition 7.32. *Suppose that X is a simplicial complex on n vertices, and suppose that we have a partition $X_0 = V_1 \cup \ldots V_t$ of the vertices of X into t classes (colors) such that the following properties hold:*

1. *The partition is an* equipartition *of the complex (of the vertices as well as of the higher-dimensional faces). That is, we assume that all $|V_i|$ are equal to $\frac{1+o(1)}{t}n$,*

and that for each dimension $i < t$ and each $A \in \binom{[t]}{i+1}$, we have $f_i(X[A]) = (1 + o(1))\left(\frac{i+1}{t}\right)^{i+1} \cdot f_i(X)$.

(This is the behavior we get with high probability when we t-color the vertices uniformly at random.)

2. *For $1 \le i \le k$ and $A \in \binom{[t]}{i+1}$, the complex $X[A]$ is coarsely face-expanding in dimension i with face expansion at least $\varepsilon_i > 0$ and coarseness $\delta_i < \frac{i!\varepsilon_1\varepsilon_2\cdots\varepsilon_{i-1}}{(i+1)^i t^i}$.*

Then there is no K_t^k-cominor for this partition; i.e., X has a colorful homological minor.

Proof. Assume that α is a cominor for the partition. By the preceding lemma, we may assume that every component α_A is cohomologically minimal.

Since $\sum_i \alpha_i(\emptyset) = 1$, we must have $\alpha_i(\emptyset) = 1$ for some i, say $\alpha_1(\emptyset) = 1$. For each $v \in V_1$, we have $\partial^*\alpha_1(v) = 1 = \sum_{i\in[t]\setminus\{1\}} \alpha_{\{1,i}\}(v)$. It follows that there is some i such that for at least $1/t$ of the vertices in V_1, we have $\alpha_{1i}(v) = 1$, say $i = 2$. It follows that $|\alpha_{12}| \ge |V_1|/t = |V_1 \cup V_2|/2t$; hence, $\|\alpha_{12}\| = \|[\alpha_{12}]\| \ge \frac{1}{2t}$ in $X[12]$. Here, we use that α_{12} is cohomologically minimal and that $|V_1| = |V_2|$.

We assume that $\delta_0 < 1/2t$ and that $X[12]$ is coarsely expanding. Thus, $\|\partial^*\alpha_{12}\| \ge \varepsilon_0/2t$ in $X[12]$. Thus, $\partial^*\alpha_{12}$ is supported by a fraction of at least $\varepsilon_1/2t$ of the edges of $X[12]$. By the cominor relation, there must be an index j such that α_{12j} is supported by a fraction of at least $\frac{\varepsilon_1}{2t^2}$ of the edges in $X[12]$, say $j = 3$.

By the equipartition property and since α_{123} is minimal, we get $\|[\alpha_{123}]\| \ge \frac{\varepsilon_1}{2t^2} \cdot \frac{4}{9}(1 + o(1))$ in $X[123]$. If $\delta_2 < \frac{2\varepsilon_1}{9t^2}$, then coarse expansion of $X[123]$ implies that $\|\partial^*\alpha_{123}\| \ge \frac{2\varepsilon_1\varepsilon_2}{9t^2}$ in $X[123]$.

Inductively, we get, for every $i \le k$, a set $A \in \binom{t}{i+1}$ such that

$$\|[\alpha_A]\| \ge \frac{i!\varepsilon_1\varepsilon_2\cdots\varepsilon_{i-1}}{(i+1)^i t^i} > \delta_i$$

and

$$\|\partial^*\alpha_A\| \ge \frac{i!\varepsilon_1\varepsilon_2\cdots\varepsilon_i}{(i+1)^i t^i}$$

in $X[A]$. For $i = k$, this is a contradiction, since in this case, we should have $\partial^*\alpha_A = 0$. □

Proof of Theorem 4.17. Let $X = X^k[n, C/n]$, where C is a constant to be determined. Fix any partition (coloring) of the vertices of X into t parts of equal size, say $[n] = V_1 \cup \ldots \cup V_t$ with $V_i = \{\frac{(i-1)n}{t} + 1, \ldots, \frac{in}{t}\}$, $1 \le i \le t$.

Observe that for this coloring and each face $A \in K_t^k$, the complex $X[A]$ is isomorphic to $X^k(n', C'/n')$, where $n' = \frac{k+1}{t}n$ and $C' = \frac{k+1}{t}C$. It follows easily that asymptotically almost surely the coloring yields an equipartition of X in the sense of Proposition 7.32.

Moreover, by Lemma 6.29, the necessary expansion properties hold asymptotically almost surely provided

$$\frac{k+1}{t}C = C' > \frac{8(k+1)\ln(2)(k+1)^k t^k}{k!}.$$

This proves the theorem. □

Remark 33. There is a strong formal similarity between the pigeonholing argument in the proof of Proposition 7.32 and the combinatorial part of Gromov's proof of the *filling-contraction inequality* in [20, Sect. 2.4]. On the other hand, the underlying topological arguments that lead to these combinatorial problems seem to be different, at least at first sight. It would be interesting to understand this connection better.

Moreover, we hope that the approach presented here will also be useful for the corresponding extremal problems. This is work in progress.

Acknowledgements The author would like to thank János Pach and Emo Welzl for the invitation to participate in the Special Semester on Discrete and Computational Geometry and for the opportunity to present my results there.

The author would also like to thank Gil Kalai, Roy Meshulam, and Eran Nevo for helpful comments and discussions.

The author's research was supported by the Swiss National Science Foundation, SNF Projects 200021-125309, 20020-125027, and 200020-138230.

References

1. M. Ajtai, V. Chvátal, M. Newborn, E. Szemerédi, Crossing-free subgraphs. Ann. Discrete Math. **12**, 9–12 (1982)
2. L. Aronshtam, N. Linial, T. Luczak, R. Meshulam, Collapsibility and vanishing of top homology in random simplicial complexes. Discrete Comput. Geom., to appear, `arXiv:1010.1400v2`, 2010
3. E. Babson, C. Hoffman, M. Kahle, The fundamental group of random 2-complexes. J. Am. Math. Soc. **24**(1), 1–28 (2011)
4. R. Bing, An alternative proof that 3-manifolds can be triangulated. Ann. Math. (2) **69**, 37–65 (1959)
5. B. Bollobás, P.A. Catlin, P. Erdős, Hadwiger's conjecture is true for almost every graph. Eur. J. Combin. **1**(3), 195–199 (1980)
6. B. Bollobás, A. Thomason, Proof of a conjecture of Mader, Erdős and Hajnal on topological complete subgraphs. Eur. J. Combin. **19**(8), 883–887 (1998)
7. U. Brehm, K.S. Sarkaria, Linear vs. piecewise linear embeddability of simplicial complexes. Techenical Report 92/52, Max-Planck-Institut für Mathematik, Bonn, Germany, 1992
8. W.G. Brown, P. Erdős, V.T. Sós, Some extremal problems on r-graphs, in *New Directions in the Theory of Graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, MI, 1971)* (Academic Press, New York, 1973), pp. 53–63
9. J. Bryant, Approximating embeddings of polyhedra in codimension 3. Trans. Am. Math. Soc. **170**, 85–95 (1972)

10. D. Cohen, A. Costa, M. Farber, T. Kappeler, Topology of random 2-complexes. Discrete Comput. Geom. **47**(1), 117–149 (2012)
11. T.K. Dey, On counting triangulations in d dimensions. Comput. Geom. **3**(6), 315–325 (1993)
12. T.K. Dey, Improved bounds on planar k-sets and related problems. Discrete Comput. Geom. **19**, 373–382 (1998)
13. T.K. Dey, H. Edelsbrunner, Counting triangle crossings and halving planes. Discrete Comput. Geom. **12**, 281–289 (1994)
14. T.K. Dey, J. Pach, Extremal problems for geometric hypergraphs. Discrete Comput. Geom. **19**(4), 473–484 (1998)
15. T.K. Dey, N.R. Shah, On the number of simplicial complexes in $\mathbf{R}^d$. Comput. Geom. **8**(5), 267–277 (1997)
16. R. Diestel, in *Graph Theory*. Graduate Texts in Mathematics, vol. 173, 3rd edn. (Springer, Berlin, 2005)
17. P. Erdős, On extremal problems of graphs and generalized graphs. Israel J. Math. **2**, 183–190 (1964)
18. A. Flores, Über die Existenz n-dimensionaler Komplexe, die nicht in den $\mathbf{R}^{2n}$ topologisch einbettbar sind. Ergeb. Math. Kolloqu. **5**, 17–24 (1933)
19. J. Geelen, B. Gerards, G. Whittle, Towards a matroid-minor structure theory, in *Combinatorics, Complexity, and Chance*. Oxford Lecture Ser. Math. Appl., vol. 34 (Oxford University Press, Oxford, 2007), pp. 72–82
20. M. Gromov, Singularities, expanders and topology of maps. Part 2: From combinatorics to topology via algebraic isoperimetry. Geom. Funct. Anal. **20**(2), 416–526 (2010)
21. B. Grünbaum, Imbeddings of simplicial complexes. Comment. Math. Helv. **44**, 502–513 (1969)
22. A. Gundert, On the complexity of embeddable simplicial complexes. Diplomarbeit, Freie Universität Berlin, 2009
23. H. Hadwiger, Über eine Klassifikation der Streckenkomplexe. Vierteljschr. Naturforsch. Ges. Zürich **88**, 133–142 (1943)
24. A. Haefliger, Plongements de variétés dans le domaine stable. Sém. Bourbaki **245** (1982)
25. S. Janson, On concentration of probability, in *Contemporary Combinatorics*. Bolyai Soc. Math. Stud., vol. 10 (János Bolyai Math. Soc., Budapest, 2002), pp. 289–301
26. T. Kaiser, Minors of simplicial complexes. Discrete Appl. Math. **157**(12), 2597–2602 (2009)
27. G. Kalai, The diameter of graphs of convex polytopes and f-vector theory, in *Applied Geometry and Discrete Mathematics*. DIMACS Ser. Discrete Math. Theoret. Comput. Sci., vol. 4 (Amer. Math. Soc., Providence, RI, 1991), pp. 387–411
28. G. Kalai, Algebraic shifting, in *Computational Commutative Algebra and Combinatorics (Osaka, 1999)*. Adv. Stud. Pure Math., vol. 33 (Math. Soc. Japan, Tokyo, 2002), pp. 121–163
29. J. Komlós, E. Szemerédi, Topological cliques in graphs. II. Combin. Probab. Comput. **5**(1), 79–90 (1996)
30. D. Kozlov, in *Combinatorial Algebraic Topology*. Algorithms and Computation in Mathematics, vol. 21 (Springer, Berlin, 2008)
31. D. Kozlov, The threshold function for vanishing of the top homology group of random d-complexes. Proc. Am. Math. Soc. **138**(12), 4517–4527 (2010)
32. M. Krivelevich, B. Sudakov, Minors in expanding graphs. Geom. Funct. Anal. **19**(1), 294–331 (2009)
33. K. Kuratowski, Sur le problème des courbes gauches en topologie. Fund. Math. **15**, 271–283 (1930)
34. F.T. Leighton, *Complexity Issues in VLSI* (MIT Press, Cambridge, 1983)
35. N. Linial, R. Meshulam, Homological connectivity of random 2-complexes. Combinatorica **26**(4), 475–487 (2006)
36. L. Lovász, Graph minor theory. Bull. Am. Math. Soc. (N.S.) **43**(1), 75–86 (2006)
37. T. Łuczak, B. Pittel, J.C. Wierman, The structure of a random graph at the point of the phase transition. Trans. Am. Math. Soc. **341**(2), 721–748 (1994)

38. W. Mader, Homomorphieeigenschaften und mittlere Kantendichte von Graphen. Math. Ann. **174**, 265–268 (1967)
39. W. Mader, $3n - 5$ edges do force a subdivision of K_5. Combinatorica **18**(4), 569–595 (1998)
40. J. Matoušek, *Using the Borsuk–Ulam Theorem* (Springer, Berlin, 2003)
41. J. Matoušek, M. Tancer, U. Wagner, Hardness of embedding simplicial complexes into $\mathbf{R}^d$, in *Proceedings of the Twentieth ACM-SIAM Symposium on Discrete Algorithms (SODA'09)*, SIAM, Philadelphia, PA, 2009, pp. 855–864
42. J. Matoušek, M. Tancer, U. Wagner, Hardness of embedding simplicial complexes in $\mathbf{R}^d$. J. Eur. Math. Soc. **13**(2), 259–295 (2011)
43. P. McMullen, The maximum numbers of faces of a convex polytope. Mathematika **17**, 179–184 (1970)
44. P. McMullen, The numbers of faces of simplicial polytopes. Israel J. Math. **9**, 559–570 (1971)
45. S.A. Melikhov, The van Kampen obstruction and its relatives. Proc. Steklov Inst. Math. **266**(1), 142–176 (2009)
46. S.A. Melikhov, Combinatorics of embeddings. Preprint, `arxiv:1103.5457v2`, 2011
47. R. Meshulam, N. Wallach, Homological connectivity of random k-dimensional complexes. Random Struct. Algorithm. **34**(3), 408–417 (2009)
48. J.R. Munkres, *Elements of Algebraic Topology* (Addison-Wesley, Menlo Park, CA, 1984)
49. E. Nevo, Higher minors and van Kampen's obstruction. Math. Scand. **101**, 161–176 (2007)
50. E. Nevo, U. Wagner, On the embeddability of skeleta of spheres. Israel J. Math. **174**, 381–402 (2009)
51. I. Newman, Y. Rabinovich, Finite volume spaces and sparsification. Preprint, `arXiv:1002.3541v3`, 2010
52. C. Papakyriakopoulos, A new proof for the invariance of the homology groups of a complex (in Greek). Bull. Soc. Math. Grèce **22**(1943), 1–154 (1946)
53. V.V. Prasolov, in *Elements of Homology Theory*. Graduate Studies in Mathematics, vol. 81 (American Mathematical Society, Providence, RI, 2007)
54. D. Repovš, A.B. Skopenkov, New results on embeddings of polyhedra and manifolds into Euclidean spaces. Uspekhi Mat. Nauk **54**(6(330)), 61–108 (1999)
55. N. Robertson, P. Seymour, R. Thomas, Sachs' linkless embedding conjecture. J. Combin. Theor. Ser. B **64**(2), 185–227 (1995)
56. N. Robertson, P.D. Seymour, Graph minors. XX. Wagner's conjecture. J. Combin. Theor. Ser. B **92**(2), 325–357 (2004)
57. N. Robertson, P.D. Seymour, R. Thomas, Linkless embeddings of graphs in 3-space. Bull. Am. Math. Soc. (N.S.) **28**(1), 84–89 (1993)
58. H. Sachs, On spatial representations of finite graphs, in *Finite and Infinite Sets, Vol. I, II (Eger, 1981)*. Colloq. Math. Soc. János Bolyai, vol. 37 (North-Holland, Amsterdam, 1984), pp. 649–662
59. K.S. Sarkaria, Shifting and embeddability of simplicial complexes. Technical report, Max-Planck-Institut für Mathematik, Bonn, MPI/92-51, 1992
60. M. Sharir, E. Welzl, Random triangulations of planar point sets, in *Proceedings of the 22nd Annual ACM Symposium on Computational Geometry (SoCG)* (ACM, New York, 2006), pp. 273–281
61. A.B. Skopenkov, Embedding and knotting of manifolds in Euclidean spaces, in *Surveys in Contemporary Mathematics*. London Math. Soc. Lecture Note Ser., vol. 347 (Cambridge University Press, Cambridge, 2008), pp. 248–342
62. R.P. Stanley, The upper bound conjecture and Cohen–Macaulay rings. Studies Appl. Math. **54**(2), 135–142 (1975)
63. R.P. Stanley, The number of faces of simplicial polytopes and spheres, in *Discrete Geometry and Convexity (New York, 1982)*. Ann. New York Acad. Sci., vol. 440 (New York Academy of Sciences, New York, 1985), pp. 212–223
64. L.A. Székely, Crossing numbers and hard Erdős problems in discrete geometry. Combin. Probab. Comput. **6**(3), 353–358 (1997)

65. A. Thomason, The extremal function for complete minors. J. Combin. Theor. Ser. B **81**(2), 318–338 (2001)
66. B.R. Ummel, Imbedding classes and *n*-minimal complexes. Proc. Am. Math. Soc. **38**, 201–206 (1973)
67. E.R. van Kampen, Komplexe in euklidischen Räumen. Abh. Math. Sem. Univ. Hamburg **9**, 72–78 (1932)
68. K. Wagner, Über eine Eigenschaft der ebenen Komplexe. Math. Ann. **114**(1), 570–590 (1937)
69. U. Wagner, Minors in random and expanding hypergraphs, in *Proceedings of the 27th Annual Symposium on Computational Geometry (SoCG)* (ACM, New York, 2011), pp. 351–360
70. C. Weber, Plongements de polyedres dans le domaine metastable. Comment. Math. Helv. **42**, 1–27 (1967)
71. J. Zaks, On minimal complexes. Pacif. J. Math. **28**, 721–727 (1969)

Erratum to:

Thirty Essays on Geometric Graph Theory*

Greg Aloupis, Brad Ballinger, Sébastien Collette, Stefan Langerman, Attila Pór, and David R. Wood

∗This is the full version of an extended abstract presented at the 26th European Workshop on Computational Geometry (EuroCG'10).

G. Aloupis • S. Collette
Chargé de Recherches du F.R.S.-FNRS, Département d'Informatique,
Université Libre de Bruxelles, Brussels, Belgium
e-mail: aloupis.greg@gmail.com; sebastien.collette@ulb.ac.be

B. Ballinger
Department of Mathematics, Humboldt State University, Arcata, CA, USA
e-mail: bjb86@humboldt.edu

S. Langerman
Maître de Recherches du F.R.S.-FNRS, Département d'Informatique,
Université Libre de Bruxelles, Brussels, Belgium
e-mail: stefan.langerman@ulb.ac.be

A. Pór
Department of Mathematics, Western Kentucky University, Bowling Green,
KY, USA
e-mail: attila.por@wku.edu

D.R. Wood (✉)
Department of Mathematics and Statistics, The University of Melbourne,
Melbourne, Australia
e-mail: woodd@unimelb.edu.au

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0, pp. 31–48,

The publisher regrets that the fourth chapter has incorrect title. The correct title is, "**Blocking Colored Point Sets**".

The online version of the original chapter can be found at
http://dx.doi.org/10.1007/978-1-4614-0110-0_4

Erratum Continued on Next Page ⟶

Erratum to:

Ramsey-Type Problems for Geometric Graphs

Gyula Károlyi

Institute of Mathematics, Eötvös University Pázmány P. sétány 1/c,
Budapest, H–1117 Hungary
e-mail: karolyi@cs.elte.hu

J. Pach (ed.), *Thirty Essays on Geometric Graph Theory*,
DOI 10.1007/978-1-4614-0110-0, pp. 371–382,

The publisher regrets the following errors in the nineteenth chapter:

On page 375, line 28, the spelling "Erőd" is incorrect. The correct spelling is "Erdős.

On page 380, line 22 should read:
"**Problem 9**. Is there a constant $\alpha > 1$ such that $RM_{\mathrm{c}}(M_{2\mathrm{n}}) = \Omega(\alpha^{\mathrm{n}})$?"

On page 380, following **Problem 10** the following section was omitted:

6 Note Added in Proof
Problem 9 was settled by József Borbély, demonstrating that $RM_{\mathrm{c}}(M_{2\mathrm{n}}) = 1$.

10.1007/978-1-4614-0110-0_32

The online version of the original chapter can be found at
http://dx.doi.org/10.1007/978-1-4614-0110-0_19