Algorithms and Combinatorics 26

Martin Klazar · Jan Kratochvíl
Martin Loebl · Jiří Matoušek
Robin Thomas · Pavel Valtr

Editors

Topics in Discrete Mathematics

Dedicated to Jarik Nešetřil on the Occasion of his 60th Birthday

With 62 Figures

Springer

Editors

Martin Klazar
Jan Kratochvíl
Martin Loebl
Jiří Matoušek
Pavel Valtr

Department of Applied Mathematics
and Institute for Theoretical
Computer Science ITI
Faculty of Mathematics and Physics
Charles University, Prague
Malostranské nám. 25
118 00 Praha 1, Czech Republic

Robin Thomas

School of Mathematics
Georgia Institute of Technology
Atlanta, GA 30332-0160, USA

Library of Congress Control Number: 2006926241

Mathematics Subject Classification (2000):
05C10, 05C15, 05C30, 05C55, 05C65, 05C70, 05C85

ISSN 0937-5511

ISBN-10 3-540-33698-2 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-33698-3 Springer Berlin Heidelberg New York

Springer is a part of Springer Science+Business Media

springer.com

Printed in Germany

Typesetting: by the authors, copyedited by Helena Nyklová using a Springer LaTeX macro package
Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig
Cover design: *design & production* GmbH, Heidelberg

Printed on acid-free paper 46/3100YL - 5 4 3 2 1 0

To our teacher, colleague and friend

Preface

The purpose of this book is twofold. We would like to offer our readers a collection of high quality papers in selected topics of Discrete Mathematics, and, at the same time, celebrate the 60th birthday of Jarik Nešetřil. Since our discipline has experienced an explosive growth during the last half century, it is impossible to cover all of its recent developments in one modest volume. Instead, we concentrate on six topics, those closest to Jarik's interests. We have invited leading experts and close friends of Jarik's to contribute to this endeavor, and the response has been overwhelmingly positive. We were fortunate to receive many outstanding contributions. They are divided into six parts.

Contents

The topics of the first part are rather diverse, including Algebra, Geometry, and Numbers and Games. Michael E. Adams and Aleš Pultr consider rigidity (lack of nontrivial homomorphisms) of algebraic structures, and they construct $2^{\aleph_0}$ rigid countable Heyting algebras. Vitaly Bergelson, Hillel Furstenberg, and Benjamin Weiss introduce a new notion of "large" sets of integers, piecewise-Bohr sets, and they show, in particular, that the sum of two sets of positive upper density is piecewise Bohr. Christopher Cunningham and Igor Kriz investigate a generalization of the Conway number games to more than two players and construct games with any given value. Miroslav Fiedler solves extremal geometric questions, namely, the shape of n-dimensional unit-volume simplices that maximize the length of a Hamilton cycle or path in their graph. The paper of Václav Koubek and Jiří Sichler in universal algebra concerns the relation of Q-universality and finite-to-finite universality of algebras. Christian Krattenthaler studies a simplicial complex associated to a colored root system, the generalized cluster complex, and proves a generalization of remarkable relations, discovered by Chapoton, concerning certain face counts.

Part II contains contributions in Ramsey theory. Ron Graham and József Solymosi give an elementary proof that an $n \times n$ integer grid colored by fewer than $\log \log n$ colors contains a monochromatic vertex set of an equilateral right triangle. András Gyárfás, Miklós Ruszinkó, Gábor N. Sárközy, and Endre Szemerédi use the regularity lemma for constructing coverings of edge-r-colored complete bipartite graphs by vertex-disjoint monochromatic cycles. Neil Hindman and Imre Leader consider a variant of partition-regularity of systems of linear equations, where they look for nonconstant solutions. Alexandr Kostochka and Naeem Sheikh construct infinitely many graphs for which the ratio of the induced Ramsey number to the weak induced Ramsey number is bounded away from 1, answering a question of Łuczak and Gorgol. Pavel Pudlák applies the recent Bourgain–Katz–Tao theorem on sums and products in finite fields to an explicit construction of 3-colorings of complete bipartite graphs with no large monochromatic complete bipartite subgraphs.

Topics in graph and hypergraph theory begin with Part III. József Balogh, Béla Bollobás, and Robert Morris consider the enumeration of ordered graphs not containing any ordered subgraph from a fixed (possibly infinite) set. The contribution of Zoltán Füredi, Kyung-Won Hwang, and Paul Weichsel is best described by its title: A proof and generalizations of the Erdős–Ko–Rado theorem using the method of linearly independent polynomials. Tomáš Kaiser, Daniel Král', and Serguei Norine prove that in any cubic bridgeless graph at least 60% of edges can be covered by two matchings, a result related to a conjecture of Berge and Fulkerson. Brendan Nagle, Vojtěch Rödl, and Mathias Schacht apply the hypergraph regularity method, a recent hypergraph generalization of the Szemerédi regularity lemma, to extremal problems for hypergraphs. Colin McDiarmid, Angelika Steger, and Dominic Welsh define addable graph classes, which include planar graphs and many other natural classes, and show that the probability of a random graph from such a class being connected is bounded away from both 0 and 1.

The papers in Part IV deal with graph homomorphisms. Noga Alon and Asaf Shapira survey the role of homomorphisms in recent results on constant-time probabilistic testing of graph properties. Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, and Katalin Vesztergombi look at the number of homomorphisms $G \to H$ from various perspectives such as graph isomorphism, reconstruction, probabilistic property testing, and statistical physics. Josep Díaz, Maria Serna, and Dimitrios Thilikos investigate an algorithmic problem, the fixed-parameter complexity of testing the existence of a homomorphism $G \to H$, where H is fixed, G is the input, and the number of preimages of certain vertices of H is restricted. Pavol Hell considers the Dichotomy Conjecture, stating that every class of constraint satisfaction problems specified by a fixed relational structure H is either polynomial-time solvable or NP-complete, establishes special cases, and connects the problem to graph colorings.

Part V is concerned mostly with generalized graph colorings. Those in the paper by Glenn Chappell, John Gimbel, and Chris Hartman are path-colorings of planar graphs. Dwight Duffus, Vojtěch Rödl, Bill Sands, and Norbert Sauer consider the minimum chromatic number of graphs and hypergraphs of large girth that cannot be homomorphically mapped to a specified graph or hypergraph, obtaining a new probabilistic hypergraph construction in the process. Mickaël Montassier, André Raspaud, and Weifan Wang prove acyclic 4-choosability of planar graphs with excluded cycles of certain lengths. Xuding Zhu presents an authoritative survey of the circular chromatic number, a parameter introduced by Vince in 1988 that carries more information than the chromatic number itself. The contribution of Claude Tardif sticks to the usual chromatic number and provides an algorithmic version of a special case of the celebrated Hedetniemi conjecture.

Part VI on graph embeddings opens with the paper by Hubert de Fraysseix and Patrice Ossona de Mendez, who consider embeddings of multigraphs in the k-dimensional Euclidean space such that automorphisms correspond to isometries and present an elegant characterization of such embeddings. Bojan Mohar extends an intriguing result of Youngs on quadrangulations of the projective plane, and constructs the first explicit family of infinitely many 5-critical graphs on a fixed surface. János Pach and Géza Tóth relate the torus crossing number of a graph to the planar crossing number. The survey of Jozef Širáň deals with the classification of regular maps (maps possessing the highest level of symmetry – their automorphism groups act transitively on the set of flags) and explains its intriguing connections to other branches of mathematics.

Presented in a part of its own comes the last article written by Jørgen Bang-Jensen, Bruce Reed, Mathias Schacht, Robert Šámal, Bjarne Toft and Ulrich Wagner about six problems posed by Jarik Nešetřil and their current status. This last paper is just a small example of the enormous influence Jarik has had on other researchers.

Dedication

Jarik Nešetřil is a scientist and an artist of extraordinary breadth and vision. His publication record and other achievements, including over a half-dozen textbooks and monographs, an honorary doctorate and an academy membership, speak for themselves. Equally important is Jarik's tireless work with students and younger colleagues. He founded the Prague Combinatorics Seminar, which helped shape the careers of several generations of Czech mathematicians and computer scientists. Among them, the present editors greatly benefited from Jarik's guidance, ideas, and endless enthusiasm. We would like to express our appreciation and wish him many more productive years filled with success and satisfaction.

Acknowledgement. Many people helped us with this volume. We are indebted to the referees, who generously gave their time and effort in order to improve the presentation of the contributions. In preparation of the final version we were greatly assisted by our technical editor Helena Nyklová, whose meticulous copyediting is warmly appreciated. We also thank Ms. J. Borkovcová for her kind permission to reprint the photograph of Jarik. Finally, we would like to thank the Institute for Theoretical Computer Science and Department of Applied Mathematics of Charles University for their support.[1]

Prague and Atlanta,
13th March 2006

Martin Klazar
Jan Kratochvíl
Martin Loebl
Jiří Matoušek
Robin Thomas
Pavel Valtr

[1] In the framework of Czech research grants 1M0021620808 and MSM0021620838.

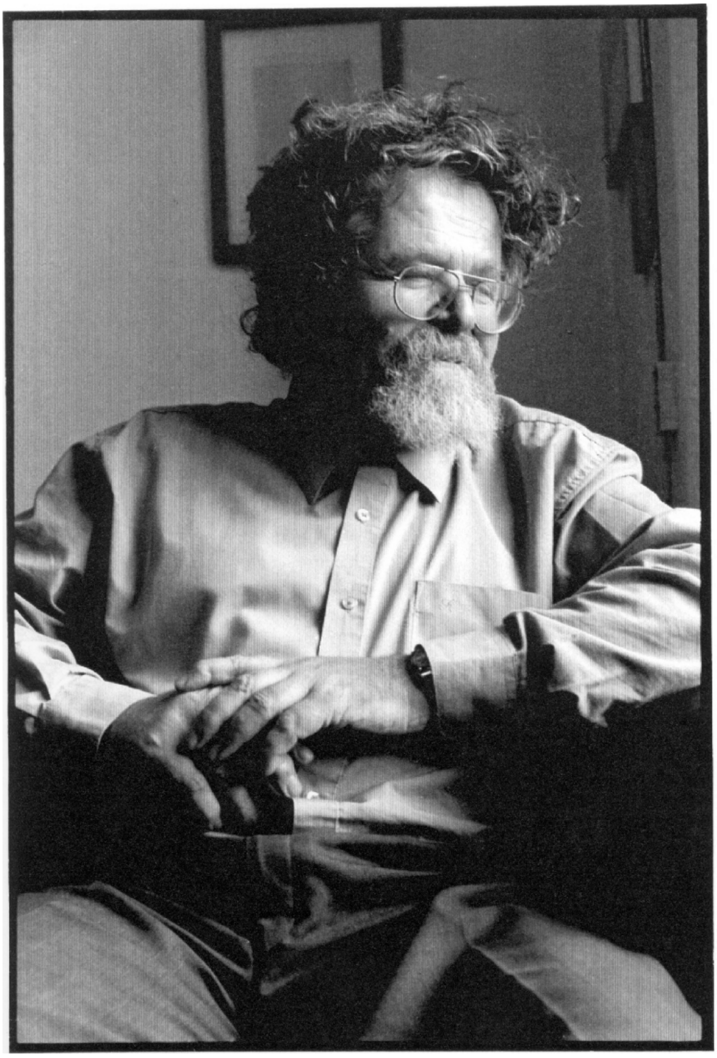


Photo of Jarik Nešetřil by Stanislav Tůma, © J. Borkovcová

Contents

Part I Algebra, Geometry, Numbers

Countable Almost Rigid Heyting Algebras
Michael E. Adams, Aleš Pultr 3

Piecewise-Bohr Sets of Integers and Combinatorial Number Theory
Vitaly Bergelson, Hillel Furstenberg, Benjamin Weiss 13

A Generalization of Conway Number Games to Multiple Players
Christopher Cunningham, Igor Kriz 39

Two Isoperimetric Problems for Euclidean Simplices
Miroslav Fiedler 65

On Finitely Generated Varieties of Distributive Double p-algebras and their Subquasivarieties
Václav Koubek, Jiří Sichler 71

The F-triangle of the Generalised Cluster Complex
Christian Krattenthaler 93

Part II Ramsey Theory

Monochromatic Equilateral Right Triangles on the Integer Grid
Ron Graham, József Solymosi 129

One-sided Coverings of Colored Complete Bipartite Graphs
András Gyárfás, Miklós Ruszinkó, Gábor N. Sárközy, Endre Szemerédi . 133

Nonconstant Monochromatic Solutions to Systems of Linear Equations
Neil Hindman, Imre Leader 145

On the Induced Ramsey Number $IR(P_3, H)$
Alexandr Kostochka, Naeem Sheikh 155

On Explicit Ramsey Graphs and Estimates of the Number of Sums and Products
Pavel Pudlák 169

Part III Graphs and Hypergraphs

Hereditary Properties of Ordered Graphs
József Balogh, Béla Bollobás, Robert Morris 179

A Proof and Generalizations of the Erdős–Ko–Rado Theorem Using the Method of Linearly Independent Polynomials
Zoltán Füredi, Kyung-Won Hwang, Paul M. Weichsel 215

Unions of Perfect Matchings in Cubic Graphs
Tomáš Kaiser, Daniel Král', Serguei Norine 225

Random Graphs from Planar and Other Addable Classes
Colin McDiarmid, Angelika Steger, Dominic J. A. Welsh 231

Extremal Hypergraph Problems and the Regularity Method
Brendan Nagle, Vojtěch Rödl, Mathias Schacht 247

Part IV Homomorphisms

Homomorphisms in Graph Property Testing
Noga Alon, Asaf Shapira 281

Counting Graph Homomorphisms
Christian Borgs, Jennifer Chayes, László Lovász, Vera T. Sós, Katalin Vesztergombi 315

Efficient Algorithms for Parameterized H-colorings
Josep Díaz, Maria Serna, Dimitrios M. Thilikos 373

From Graph Colouring to Constraint Satisfaction: There and Back Again
Pavol Hell 407

Part V Graph Colorings

Thresholds for Path Colorings of Planar Graphs
Glenn G. Chappell, John Gimbel, Chris Hartman 435

Chromatic Numbers and Homomorphisms of Large Girth Hypergraphs
Dwight Duffus, Vojtěch Rödl, Bill Sands, Norbert Sauer 455

Acyclic 4-Choosability of Planar Graphs Without Cycles of Specific Lengths
Mickaël Montassier, André Raspaud, Weifan Wang 473

On the Algorithmic Aspects of Hedetniemi's Conjecture
Claude Tardif 493

Recent Developments in Circular Colouring of Graphs
Xuding Zhu 497

Part VI Graph Embeddings

Regular Embeddings of Multigraphs
Hubert de Fraysseix, Patrice Ossona de Mendez 553

Quadrangulations and 5-critical Graphs on the Projective Plane
Bojan Mohar 565

Crossing Number of Toroidal Graphs
János Pach, Géza Tóth 581

Regular Maps on a Given Surface: A Survey
Jozef Širáň 591

Part VII

On Six Problems Posed by Jarik Nešetřil
Jørgen Bang-Jensen, Bruce Reed, Mathias Schacht, Robert Šámal, Bjarne Toft, Uli Wagner 613

Part I

Algebra, Geometry, Numbers

Countable Almost Rigid Heyting Algebras

Michael E. Adams[1] and Aleš Pultr[2]*

[1] Department of Mathematics, State University of New York, New Paltz, NY 12561, USA
adamsm@newpaltz.edu

[2] Department of Applied Mathematics and ITI, MFF, Charles University, Malostranské nám. 25, CZ 11800 Praha 1, Czech Republic
pultr@kam.ms.mff.uni.cz

Summary. For non-trivial Heyting algebras H_1, H_2 one always has at least one homomorphism $H_1 \to H_2$; if $H_1 = H_2$ there is at least one non-identical one. A Heyting algebra H is *almost rigid* if $|\mathsf{End}(H)| = 2$ and a system of almost rigid algebras $\mathcal{H}$ is said to be *discrete* if $|\mathsf{Hom}(H_1, H_2)| = 1$ for any two distinct $H_1, H_2 \in \mathcal{H}$. We show that there exists a discrete system of 2^ω countable almost rigid Heyting algebras.

AMS Subject Classification. 06D20, 18A20, 18B15.
Keywords. Heyting algebras, almost rigid, discrete system, Priestley duality.

Introduction

A *Heyting algebra*

$$(H; \vee, \wedge, \to, 0, 1)$$

is an algebra of type $(2,2,2,0,0)$ where $(H; \vee, \wedge, 0, 1)$ is a distributive $(0,1)$-lattice and the extra operation $x \to y$ satisfies the formula

$$z \leq x \to y \quad \text{iff} \quad x \wedge z \leq y.$$

That is to say, a Heyting algebra is a bounded relatively pseudocomplemented distributive lattice for which relative pseudocomplementation is taken to be a binary algebraic operation.

Since any two elements of a finite distributive lattice have a uniquely determined relative pseudocomplement, any finite distributive lattice can be viewed as a Heyting algebra. So too, any two elements of a Boolean algebra

* The second author would like to express his thanks for the support by the project LN 00A056 of the Ministry of Education of the Czech Republic.

have a uniquely determined relative pseudocomplement $a \to b = \neg a \vee b$ which also provides an example of a Heyting algebra.

For an algebra A, let $\mathsf{Aut}(A)$ (resp $\mathsf{End}(A)$) denote the group of automorphisms (resp. the monoid of endomorphisms) of A under the operation of composition. An algebra is *automorphism rigid* provided $|\mathsf{Aut}(A)| = 1$, that is, the only automorphism of A is the identity.

Independently, Jónsson [Jón51], Katětov [Kat51], Kuratowski [Kur26], and Rieger [Rie51] have shown that there exists a proper class of non-isomorphic automorphism rigid Boolean algebras. Since, as observed above, every Boolean algebra is relatively pseudocomplemented, there exists a proper class of non-isomorphic automorphism rigid Heyting algebras as well.

In sharp contrast with respect to their automorphism groups, as independently shown by Magill [Mag72], Maxson [Max72], and Schein [Sch70], Boolean algebras are uniquely recoverable from their endomorphism monoids. That is, for Boolean algebras B and B', if $\mathsf{End}(B) \cong \mathsf{End}(B')$, then $B \cong B'$; or, by the result of Tsinakis ([Tsi79]), for bounded relative Stone lattices which are principal, $\mathsf{End}(S) \cong \mathsf{End}(S')$ implies $S \cong S'$ as well. However, this is far from the case for general Heyting algebras, where endomorphisms can be very few.

There are necessary non-identical homomorphisms, though. Every non-trivial Heyting algebra has at least one minimal prime ideal. Furthermore, for each minimal prime ideal I of a Heyting algebra H, and any other Heyting algebra H', $\varphi(x) = 0$ for $x \in I$ and 1 otherwise determines an homomorphism $\varphi : H \to H'$. Such homomorphisms will be referred to as

trivial homomorphisms.

In particular, if $|H| \geq 3$, then $|\mathsf{End}(H)| \geq 2$. Dismissing the necessary trivial endomorphisms one defines an *almost rigid* Heyting algebra H as such that $|\mathsf{End}(H)| = 2$. Thus, by the preceding remarks, H is almost rigid if and only if $|H| \geq 3$, if H has exactly one minimal prime ideal and the only endomorphism other than the identity is associated with the minimal prime as indicated.

In [AKS85] it was shown that there exists a proper class of non-isomorphic almost rigid Heyting algebras. All the almost rigid Heyting algebras from [AKS85] have cardinality $\geq 2^\omega$. Taking into account that for $|H| \geq 4$ there is no almost rigid finite Heyting algebra (for any such H either there are at least two minimal prime ideals I and I' or else a minimal prime ideal I and another prime ideal I' which is minimal with respect to properly containing I; in the former case obviously $|\mathsf{End}(H)| \geq 3$, in the latter case, for $a \in I' \setminus I$, $\psi(x) = 0$ for $x \in I$, a for $x \in I' \setminus I$, and 1 otherwise determines an endomorphism $\psi \in \mathsf{End}(H)$ distinct from φ associated with I, and $|\mathsf{End}(H)| \geq 3$ again) this begs the question whether there are countable almost rigid Heyting algebras. This is answered in the positive in this article. Moreover, we show that

there exists a system $\mathcal{H}$ of 2^ω countable almost rigid Heyting algebras such that

- *for each A in $\mathcal{H}$, $\mathsf{End}(H) = \{\mathrm{id}_A, \tau_{AA}\}$, and*

– *for any two distinct A, B in $\mathcal{H}$ there is exactly one homomorphism τ_{AB} : $A \to B$,*

where the τ_{AB} are unique trivial homomorphisms.

For related background on Heyting or Boolean algebras see Balbes and Dwinger [BD74] or Koppelberg [Kop89].

1 Preliminaries

Let $(P, \leq)$ be a partially ordered set. For $Q \subseteq P$, set $\downarrow Q = \{x \in P \mid x \leq y \text{ for some } y \in Q\}$ and $\uparrow Q = \{x \in P \mid x \geq y \text{ for some } y \in Q\}$; for $Q = \{x\}$ we write just $\downarrow x$ and $\uparrow x$, respectively. A set Q is said to be *decreasing* or *increasing* if $Q = \downarrow Q$ or $Q = \uparrow Q$, respectively. For partially ordered sets P and P', a mapping $\varphi : P \longrightarrow P'$ is *order-preserving* providing $\varphi(x) \leq \varphi(y)$ whenever $x \leq y$.

A *Priestley space* $(P; \leq, \tau)$ is a partially ordered set $(P, \leq)$ endowed with a compact topology τ which is totally order-disconnected (namely, for any $x, y \in P$ such that $x \not\leq y$ there exists a clopen decreasing set $Q \subseteq P$ such that $x \notin Q$ and $y \in Q$).

As shown by Priestley [Pri70], [Pri72], the category of non-trivial distributive $(0,1)$-lattices together with all $(0,1)$-lattice homomorphisms is dually isomorphic to the category of Priestley spaces together with all continuous order-preserving maps. The equivalence functors are usually given as

$$\mathcal{P}(L) = \{x \mid L \neq x \text{ a prime ideal of } L\}, \quad \mathcal{P}(h)(x) = h^{-1}[x],$$
$$\mathcal{D}(X) = \{U \mid U = \downarrow U \subseteq X \text{ clopen }\}, \quad \mathcal{D}(f)(U) = f^{-1}[U];$$

$\mathcal{P}(L)$ is endowed with a suitable topology and ordered by inclusion.

Since every Heyting algebra is a distributive $(0,1)$-lattice, it is to be expected that the category of all non-trivial Heyting algebras is dually isomorphic to a well-defined subcategory of the category of all Priestley spaces. And indeed, the Priestley spaces X dual to Heyting algebras are precisely those with the additional property that $\uparrow U$ is clopen whenever U is clopen. Such Priestley spaces will be called

h-spaces,

and if L, M are Heyting algebras then the Heyting homomorphisms $h : L \to M$ correspond to the Priestley maps f such that, moreover,

$$f(\downarrow x) = \downarrow f(x).$$

Such maps will be referred to as

h-maps.

It is this dual equivalence that we will use in order to establish our result.

2 The Construction

The Posets

Set $X = \{n \in \mathbb{N} \mid n \geq 5\}$ and decompose this set as follows. Start with

$$X_1 = \{5\}, \quad \phi(2) = 5$$
$$X_2 = X_{2,1} = \{6, 7, 8, 9, 10\}, \quad \phi(3) = 10,$$

and further proceed inductively: if $X_k = \{\phi(k)+1, \phi(k)+2, \ldots, \phi(k+1)\}$ is already defined (and, hence, $\phi(k)$ and $\phi(k+1)$ too), take, for each element $\phi(k)+j \in X_k$, a set $X_{k+1,j}$ determined as follows

$X_{k+1,1} = \{\phi(k+1)+1, \ldots, \phi(k+1)+\phi(k)+1\}$, the first $\phi(k)+1$ natural numbers after $\phi(k+1)$,
$X_{k+1,2} = \{\phi(k+1)+\phi(k)+2, \phi(k+1)+\phi(k)+3, \ldots, \phi(k+1)+2\phi(k)+3\}$, the next $\phi(k)+2$ natural numbers after $\phi(k+1)+\phi(k)+1$,
where, in general, for $1 \leq j \leq \phi(k+1)-\phi(k)$,
$X_{k+1,j} = \{\phi(k+1)+(j-1)\phi(k)+\binom{j}{2}+1, \ldots, \phi(k+1)+j\phi(k)+\binom{j+1}{2}\}$, the next $\phi(k)+j$ natural numbers after $\phi(k+1)+(j-1)\phi(k)+\binom{j}{2}$.

Then set

$$X_{k+1} = \{\phi(k+1)+1, \phi(k+1)+2, \ldots, \phi(k+2)\} = \bigcup_{j=1}^{\phi(k+1)-\phi(k)} X_{k+1,j}.$$

For triples x, y, z of distinct elements belonging to the same X_k choose distinct

$$\tau(x, y, z) \notin X$$

and set

$$T = \{\tau(x, y, z) \mid x, y, z\}$$

and

$$Y = X \cup T \cup \{\omega\}$$

where ω is an element $\notin X \cup T$.

Now choose a countably infinite system $\mathbb{Q}$ of quadruples $\{x_1, x_2, x_3, x_4\}$ such that

- for every $q = \{x_1, x_2, x_3, x_4\}$, $q \subseteq X_k$ for some k, and
- if $p, q \in \mathbb{Q}$, $p \neq q$, then $p \cap q = \emptyset$.

For $A \subseteq \mathbb{Q}$ set

$$Z(A) = Y \cup A$$

and define an order $\sqsubseteq$ on $Z(A)$ by

$$\omega \sqsubseteq x \text{ for all } x \in Z(A),$$

and by transitivity from the successor relation $\prec$ where

$$x, y, z \prec \tau(x, y, z),$$
$$x_i \prec \{x_1, x_2, x_3, x_4\} \text{ for } \{x_1, x_2, x_3, x_4\} \in A,$$
$$\text{and for } x \in X_{k+1,j}, \quad x \prec \phi(k) + j.$$

Note that

$$\uparrow\! x \textit{ is finite for all } x \in Z(A) \setminus \{\omega\}.$$

To simplify the notation define the *degree*

$$d\,(n) = n \quad \text{for} \quad n \in X,$$
$$d\,(\tau(x, y, z)) = 3,$$
$$d\,(q) = 4 \text{ for } q \in A;$$

for ω the degree is undefined.

The Topology

$Z(A)$ is endowed with the topology in which
U is open iff either $\omega \notin U$ or $Z(A) \setminus U$ is finite.
Thus we have

Observation 2.1. *The clopen sets are precisely the finite M not containing ω and the complements of such sets,and hence the $\uparrow\! M$ with clopen M are clopen.*
If x is not $\sqsubseteq y$ then $y \notin\uparrow\! x$ and $\uparrow\! x$ is clopen. Thus,
each $Z(A)$ with the order $\sqsubseteq$ and the topology just defined is an h-space.

From this we immediately obtain

Fact 2.2. *All the Heyting algebras $\mathcal{D}(Z(A))$ are countable.*

In the sequel, f will always be an h-map $Z(A) \to Z(B)$.

Lemma 2.3. *1. $f(\omega) = \omega$.*
2. If $f(M) = \{x\}$ for an infinite $M \subseteq Z(A)$ then $x = \omega$.

Proof. 1. $\{f(\omega)\} = f(\downarrow\!\omega) = \downarrow\! f(\omega)$; hence $\downarrow\! f(\omega)$ has just one element.
2. For an infinite M we have $\omega \in \overline{M}$. Thus, $\omega = f(\omega) \in f(\overline{M}) \subseteq \overline{f(M)} = \{x\}$. ◻

A *branch* of an element $x \in Z(A)$ is any $\downarrow\! y$ with $y \prec x$. Note that the degree $d(x)$ defined above is the number of distinct branches of x.

Lemma 2.4. *1. If $t = \tau(x_1, x_2, x_3)$ (resp. $q = \{x_1, x_2, x_3, x_4\} \in A$) and $f(x_i) \sqsubset f(t)$ (resp. $f(q)$) for all i then we cannot have $d(f(t)) \geq 4$ (resp. $d(f(q)) \geq 5$).*

2. *If* $t = \tau(x_1, x_2, x_3)$ *and* $f(x_i) \sqsubset f(t)$ *for all* i *then we cannot have two* $f(x_i), f(x_j)$, $i \neq j$ *in the same branch* $\downarrow y$ *of* $f(t)$.
3. *If* $a_1, a_2, a_3 \prec a \in Z(A)$ *are distinct,* $f(a_1) \sqsubseteq f(a_2) \sqsubset f(a)$ *and* $f(a_3) \sqsubset f(a)$ *then all the* $f(a_i)$ *are in the same branch of* $f(a)$.

Proof. 1. $\downarrow f(t) = f(\downarrow t) = \{f(t)\} \cup \downarrow f(x_1) \cup \downarrow f(x_2) \cup \downarrow f(x_3)$ (resp. $\{f(q)\} \cup \downarrow f(x_1) \cup \downarrow f(x_2) \cup \downarrow f(x_3) \cup \downarrow f(x_4)$) and hence $\downarrow f(t)$ cannot have more than three (resp. four) branches.

2. $\downarrow f(t)$ consists of at least three branches and hence it cannot be covered by $\downarrow y$ and just one more branch.

3. By 2, $a \in X \cup A$. Consider $t = \tau(a_1, a_2, a_3)$. By 2, $f(t) = f(a_2)$ or $f(t) = f(a_3)$. Then either $f(a_3) \sqsubseteq f(a_2)$ or $f(a_2) \sqsubseteq f(a_3)$. □

Observation 2.5. *The map*

$$const_\omega = const_\omega^{AB} : Z(A) \to Z(B)$$

defined by $const_\omega(x) = \omega$ *for all* $x \in Z(A)$ *is an h-map.*

(This is the Priestley image of the unique trivial homomorphism between the corresponding Heyting algebras, each of which has precisely one minimal prime ideal.)

3 The Result

Lemma 3.1. *For* $a \neq \omega$ *such that* $f(a) \neq \omega$ *one cannot have* $d(f(a)) < d(a)$.

Proof. If $f(b) \sqsubset f(a)$ for all $b \prec a$ we are led to a contradiction by 2.4.3: let C consist of all the $c \prec f(a)$. Then there have to be two distinct b_1, b_2 with $f(b_i) \sqsubseteq c$ for some $c \in C$ and we have an $x \sqsubset a$ such that $c = f(x)$. Now $x \sqsubseteq b$ for some $b \prec a$ and we have $f(b_1), f(b_2) \sqsubseteq f(b)$ (of course, b can be one of the b_i). Now by 2.4.3 all the $f(b)$ with $b \prec a$ are in the same branch of $f(a)$, a contradiction.

Thus, there is an $a_1 \prec a_0 = a$ such that $f(a_1) = f(a)$ and as $d(a_1) > d(a_0)$ we can repeat the procedure to obtain

$$a_0 \succ a_1 \succ a_2 \succ \cdots \quad \text{with } f(a_i) = f(a),$$

and by 2.3.2 we have $f(a) = \omega$ contradicting the assumption. □

Thus in particular

$$f[X \cup \{\omega\}] \subseteq X \cup \{\omega\}.$$

In the following four lemmas, f is, as before, an h-map $Z(A) \to Z(B)$, but since all the facts are relevant for the restriction $X \cup \{\omega\} \to X \cup \{\omega\}$ only (and since $\downarrow (X \cup \{\omega\}) = X \cup \{\omega\}$), we can use expressions such as $f(a) = a$, $f(f(a))$, or $f(\downarrow a) = \downarrow f(a)$.

Lemma 3.2. *If $a \in X$ and $f(a) \neq \omega$ then $f(a) \sqsubseteq a$ and $f(f(a)) = f(a)$. If $f(a) \sqsubset a$, we have an $a' \sqsubseteq a$ such that $a' \succ f(a) = f(a')$.*

Proof. We already know that $d(f(a)) \geq d(a)$, hence if $f(a) \neq \omega$ we have $f(a) \in X$. Suppose that $d(f(a)) > d(a)$. Then we cannot have $f(b) \sqsubset f(a)$ for all $b \prec a$ since in such a case

$$f(\downarrow a) = \{f(a)\} \cup \bigcup \{f(\downarrow b) \mid b \prec a\}$$

cannot cover $\downarrow f(a)$.

Hence for some $a_1 \prec a$ we have $f(a_1) = f(a)$. Now $d(a_1) > d(a)$. If we still have $d(f(a)) > d(a_1)$, we can repeat the argument and ultimately we obtain $a = a_0 \succ a_1 \succ \cdots \succ a_k$ with $f(a_i) = f(a)$, $d(a_i) < d(f(a))$ for $i < k$, and $d(a_k) = d(f(a))$ (by 3.1 we cannot have $d(a_k) > d(f(a))$). Since $d(a_k) \geq 5$, a_k and $f(a) = f(a_k)$ are in X and hence $a_k = f(a_k)$ by the equality of the degrees. □

Lemma 3.3. *If for an $a \in X$ one has $f(a) = a$ then f is identical on the whole of $\downarrow a$.*

Proof. Let $x \sqsubset a$ be an element with the shortest path $x \prec x_1 \prec \cdots \prec a$ such that $f(x) \sqsubset x$. As $\downarrow a = f(\downarrow a)$, $x = f(y) \sqsubset y$ for some $y \sqsubset a$. But then y is one of the x_i which is a contradiction, by Lemma 3.2. □

Lemma 3.4. *If there is an $x \in X$ such that $f(x) = b \sqsubset x$ then there is a $y \in X$ such that $f(y) = u \sqsubset y$ and b, u are incomparable.*

Proof. $f(b) = b$ and hence f is identical on $\downarrow b$. Choose $b_1 \neq b_2$, $b_i \prec b$; thus $f(b_i) = b_i$. Let $b_1, b_2 \in X_k$. Choose a $y \in X_k$ incomparable with b (since $b \sqsubset x \sqsubseteq 5$ there exist such incomparable elements). Set

$$t = \tau(b_1, b_2, y).$$

Now $b_i = f(b_i) \sqsubset f(t)$ (we cannot have an equality as b_i are incomparable) and hence $f(y) \neq \omega$ (else $f(\downarrow t) = \{f(t)\} \cup \downarrow b_1 \cup \downarrow b_2 \neq \downarrow f(t)$). Thus, $f(y) \sqsubseteq y$. We cannot have $f(y) = y$ for all such y: in such a case f would be identical on the whole of X_k which would fix all the elements above as well, including 5 and x, contradicting the assumption (if $f(a) = a$ for all $a \prec n \in X$ then $f(n) \sqsupseteq a$ for all $a \prec n$ and hence $f(n) \sqsupseteq n$; we cannot have $f(n) \sqsupset n$ though, since that would imply $d(f(n)) < d(n)$).

Thus there has to be some such y with $u = f(y) \sqsubset y$, and now u is incomparable with b. □

Lemma 3.5. *For $\omega \neq x \in X$ one cannot have $f(x) \neq \omega$ and $d(f(x)) > d(x)$.*

Proof. Suppose there is such an x. By 3.2 and 3.4 we can choose an instance of $b \prec a$ such that $f(a) = f(b) = b$ and that there exists a $u \in X$ incomparable with b such that $f(u) = u$ is in an X_l with $l \leq k$ where b is in X_k (this can be achieved but exchanging the b and u in 3.4 if necessary). Consider a $c \prec a$, $c \neq b$ and a general $z \neq b, c$ in X_k. Set $t = \tau(b, c, z)$. Since $f(c) \sqsubseteq f(a) = f(b)$ we cannot have (see 2.4.2)

$$f(b), f(c), f(z) \sqsubset f(t).$$

Now $f(t)$ cannot be equal to $f(z)$ and distinct from the others since then $b = f(b) \sqsubset f(t) = f(z)$ and hence, z being in the same X_k as b, $d(f(z)) < d(z)$ contradicting 3.1. Thus we have $f(t)$ equal to either $f(c)$ or $f(b)$ and hence $f(z) \sqsubseteq f(b) = b$.

Thus, $f(X_k) \subseteq\downarrow b$. Take a $v \sqsubseteq u$ in X_k. Then $f(v) = v$ by 3.3 and we have a contradiction: v cannot be in $\downarrow b$ since u and b are incomparable and the subposet $(X, \sqsubseteq)$ of $Z(A)$ is a tree. ☐

Lemma 3.6. *Let $f : Z(A) \to Z(B)$ be an h-map. Then either $f(X) = \{\omega\}$ or $f(n) = n$ for all $n \in X$.*

Proof. By 3.5 and 3.1, $f(5) = \omega$ or $f(5) = 5$. Hence $f(5) = 5$ and, by 3.3, f is identical on $X =\downarrow 5$. ☐

Theorem 3.7. *Let $f : Z(A) \to Z(B)$ be an h-map. Then either f is* const_ω *or it is the inclusion map $Z(A) \subseteq Z(B)$.*

On the other hand, any inclusion $A \subseteq B$, with $A, B \subseteq \mathbb{Q}$ can be extended to an inclusion h-map $Z(A) \subseteq Z(B)$.

Proof. If $f(5) = \omega$ then $f(X) = \{\omega\}$ by monotonicity. Now if for $y \in T$ or $y \in A$ one should have $f(y) = x \neq \omega$ we had $\downarrow x \neq f(\downarrow y) = \{f(y)\} \cup \{\omega\}$. Thus, also $f(y) = \omega$.

If $f(5) \neq \omega$ then f is identical on X. This also fixes T, since for $t = \tau(x, y, z)$, $x, y, z \sqsubseteq f(t)$ and equality is impossible since x, y, z are incomparable, and since any other element greater then all the x, y, z has too many branches to be covered by $f(\downarrow t)$. Finally for $q = \{x_1, x_2, x_3, x_4\} \in A$ one cannot have $f(q) \in T$ by 3.1. Since the x_i are incomparable, $f(q)$ coincides with none of the $f(x_i) = x_i$ and hence, by 2.4.1, $f(q) \notin X$. By monotonicity, $f(q) \neq \omega$ and hence $f(q) \in B$. But there is only one $p \in \mathbb{Q}$ such that $x_i \sqsubset p$ for $i = 1, 2, 3, 4$, namely q itself. The second statement is obvious. ☐

As above, denote by $\mathbb{N}$ the set of all natural numbers.

Theorem 3.8. *There exist countable almost rigid Heyting algebras $H(A)$ associated with the subsets $A \subseteq \mathbb{N}$ such that*

- *if $A \nsubseteq B$ there is no non-trivial homomorphism $H(A) \to H(B)$, and*
- *if $A \subseteq B$ there exists exactly one non-trivial homomorphism $H(A) \to H(B)$.*

Consequently there exist 2^ω countable almost rigid Heyting algebras such that the only homomorphism between any two distinct of them is the trivial one.

Proof. The first part immediately follows from 3.7 and 2.2.

For the second statement it suffices to observe that there are 2^ω many subsets of $\mathbb{N}$ such that no two of them are in inclusion.

For any $N \subseteq \mathbb{N}$ consider the set

$$\widetilde{N} = \{2n \mid n \in N\} \cup \{2n+1 \mid n \notin N\}.$$

Then $\widetilde{N}_1 \subseteq \widetilde{N}_2$ only if $N_1 = N_2$. □

References

[AKS85] M. E. Adams, V. Koubek, and J. Sichler, *Endomorphisms and homomorphisms of Heyting algebras*, Algebra Universalis 20 (1985), 167–178.

[BD74] R. Balbes, P. Dwinger, *"Distributive Lattices"*, University of Missouri Press, Columbia, Missouri, 1974.

[Jón51] B. Jónsson, *A Boolean algebra without proper automorphisms*, Proc. Amer. Math. Soc. 2 (1951), 766–770.

[Kat51] M. Katětov, *Remarks on Boolean algebras*, Colloq. Math. 2 (1951), 229–235.

[Kop89] S. Koppelberg, *"Handbook of Boolean Algebras I"*, North-Holland, Amsterdam, 1989.

[Kur26] C. Kuratowski, *Sur la puissance de l'ensemble des "nombres de dimensions" de M. Fréchet*, Fund. Math. 8 (1926), 201–208.

[Mag72] K. D. Magill, *The semigroup of endomorphisms of a Boolean ring*, Semigroup Forum 4 (1972), 411–416.

[Max72] C. J. Maxson, *On semigroups of Boolean ring endomorphisms*, Semigroup Forum 4 (1972), 78–82.

[Pri70] H. A. Priestley, *Representation of distributive lattices by means of ordered Stone spaces*, Bull. London Math. Soc. 2 (1970), 186–190.

[Pri72] H. A. Priestley, *Ordered topological spaces and the representation of distributive lattices*, Proc. London Math. Soc. **324** (1972), 507–530.

[Rie51] L. Rieger, *Some remarks on automorphisms of Boolean algebras*, Fund. Math. 38 (1951), 209–216.

[Sch70] B. M. Schein, *Ordered sets, semilattices, distributive lattices, and Boolean algebras with homomorphic endomorphism semigroups*, Fund. Math. 68 (1970), 31–50.

[Tsi79] C. Tsinakis, *Brouwerian Semilattices Determined by their Endomorphism Semigroups*, Houston J. Math. 5 (1979), 427–436.

Piecewise-Bohr Sets of Integers and Combinatorial Number Theory

Vitaly Bergelson[1]*, Hillel Furstenberg[2], and Benjamin Weiss[3]

[1] Department of Mathematics, Ohio State University, Columbus, OH 43210, USA
vitaly@math.ohio-state.edu
[2] Institute of Mathematics, Hebrew University of Jerusalem, Jerusalem, Israel
harry@math.huji.ac.il
[3] Institute of Mathematics, Hebrew University of Jerusalem, Jerusalem, Israel
weiss@math.huji.ac.il

Summary. We use ergodic-theoretical tools to study various notions of "large" sets of integers which naturally arise in theory of almost periodic functions, combinatorial number theory, and dynamics. Call a subset of **N** a Bohr set if it corresponds to an open subset in the Bohr compactification, and a piecewise Bohr set (PWB) if it contains arbitrarily large intervals of a fixed Bohr set. For example, we link the notion of PWB-sets to sets of the form A+B, where A and B are sets of integers having positive upper Banach density and obtain the following sharpening of a recent result of Renling Jin.

Theorem. If A and B are sets of integers having positive upper Banach density, the sum set A+B is PWB-set.

AMS Subject Classification. 37A45.
Keywords. Bohr sets, Banach density, measure preserving system, Kronecker factor.

1 Introduction to Some Large Sets of Integers

In combinatorial number theory, as well as in dynamics, various notions of "large" sets arise. Some familiar notions are those of sets of positive (upper) density, syndetic sets, thick sets (also called "replete"), return-time sets (in dynamics), sets of recurrence (also known as Poincaré sets), (finite or infinite) difference sets, and Bohr sets. We will here introduce the notion of "piecewise-Bohr" sets (or PWB-sets), as well as "piecewise-Bohr$_0$" sets (or PWB$_0$-sets), and we'll show how they arise in some combinatorial number-theoretic questions.

We begin with some basic definitions and elementary considerations. We'll say that a subset $A \subset \mathbb{Z}$ has *positive upper* (Banach) *density*, $d^*(A) > 0$,

* V. Bergelson was partially supported by NSF, grant DMS-0345350.

if for some $\delta > 0$, there exist arbitrarily large intervals of integers $J = \{a, a+1, \ldots, a+l-1\}$ with $\frac{|J \cap A|}{|J|} \geq \delta$. (Here $|S|$ is the cardinality of the set S; $d^*(A) = \text{l.u.b.}\{\delta \text{ as above}\}$.) Syndetic sets are special cases of sets with positive upper density. Namely, A is *syndetic* if for some l, every interval J of integers with $|J| \geq l$ intersects A. Clearly $d^*(A) \geq 1/l$ in this case. We'll say a set A is *thick* if it contains arbitrarily long intervals; thus A is syndetic $\Leftrightarrow \mathbb{Z}\backslash A$ is not thick $\Leftrightarrow A \cap B \neq \emptyset$ for any thick set B. For any distinct r integers $\{a_1, a_2, \ldots, a_r\}$ the set $\{a_j - a_i | 1 \leq i < j \leq r\}$ is called an r-*difference* set or a Δ_r-set. Every thick set contains some r-difference set for every r. This is obvious for $r = 2$, and inductively, if A is thick and if A contains the $(r-1)$-difference set formed from $\{a_1, \ldots, a_{r-1}\}$, by choosing a_r in the middle of a large enough interval in A, we can complete this to an r-difference set. It follows that for any r, a set that meets every r-difference set is syndetic. An example of this is the set of (non-zero) differences $A - A = \{x - y : x, y \in A, x \neq y\}$ when A has positive upper density. For if $d^*(A) > 1/r$ and if the numbers $a_1, a_2, \ldots, a_r$ are distinct, the sets $A + a_1, A + a_2, \ldots, A + a_r$ cannot be disjoint; so, for some $1 \leq i < j \leq r$, $a_j - a_i \in A - A$. One conclusion which is behind much of our subsequent discussion is that if A has positive upper density, then $A - A$ is syndetic. We shall see in §3 that $d^*(A) > 0$ implies that $A - A$ is a *piecewise-Bohr* set.

Definition 1.1. *$S \subset \mathbb{Z}$ is a* Bohr *set if there exists a trigonometric polynomial $\psi(t) = \sum_{k=1}^{m} c_k e^{i\lambda_k t}$, with the λ_k real numbers, such that the set*

$$S' = \{n \in \mathbb{Z} : \operatorname{Re} \psi(n) > 0\}$$

is non-empty and $S \supset S'$. When $\psi(0) > 0$ we say S is a Bohr$_0$ set. (Compare with [Bilu97]).

The fact that a Bohr set is syndetic is a consequence of the almost periodicity of trigonometric polynomials. It is also a consequence of the "uniform recurrence" of the Kronecker dynamical system on the m-torus

$$(\theta_1, \theta_2, \ldots \theta_m) \longrightarrow (\theta_1 + \lambda_1, \theta_2 + \lambda_2, \ldots, \theta_m + \lambda_m).$$

Indeed, it is not hard to see that a set $S \subset \mathbb{Z}$ is Bohr if and only if there exist $m \in \mathbb{N}$, $\alpha \in \mathbb{T}^m$ and an open set $U \subset \mathbb{T}^m$ such that $S \supset \left\{n \in \mathbb{Z} : n\alpha \in U\right\}$.

Alternatively we can define Bohr sets and Bohr$_0$ sets in terms of the topology induced on the integers $\mathbb{Z}$ by imbedding $\mathbb{Z}$ in its Bohr compactification. Namely, a set in $\mathbb{Z}$ is Bohr if it contains an open set in the induced topology, and it is Bohr$_0$ if it contains a neighborhood of 0 in this topology.

We can apply the foregoing observations regarding $A - A$ to dynamical systems. We shall be concerned with *measure preserving systems* $(X, \mathcal{B}, \mu, T)$, where $(X, \mathcal{B}, \mu)$ is a probability space, $T: X \to X$ a measurable measure preserving transformation. We assume (for simplicity) that the system is *ergodic*

($T^{-1}A = A$ for $A \in \mathcal{B} \Rightarrow \mu(A)\mu(X\backslash A) = 0$). The ergodic theorem then ensures that for $A \in \mathcal{B}$ with $\mu(A) > 0$, the orbit $\{T^n x\}_{n\in\mathbb{Z}}$ of almost every x visits A along a set of times $V(x, A) = \{n : T^n x \in A\}$ of positive density. If we set $R_1(A) = \{n : A \cap T^{-n}A \neq \emptyset\}$ (the return time set of A), then for any x, $R_1(A) \supset V(x, A) - V(x, A)$. Hence $R_1(A)$ is syndetic. We can define a smaller set $R(A) = \{n : \mu(A \cap T^{-n}A) > 0\} = R(A')$ where $A' = A \setminus \bigcup\{(A \cap T^{-n}A) : \mu(A \cap T^{-n}A) = 0\}$, and it follows that $R(A)$ is also syndetic. This can be seen directly as well (and for arbitrary measure preserving systems), but the present argument illustrates the connection of dynamics to combinatorial properties of sets. We shall call sets containing sets of the form $R(A)$, where $\mu(A) > 0, RT$-sets (for return time). A set meeting every RT-set is called a Poincaré set since Poincaré's recurrence theorem gives content to the property by implying that $R(A)$ is never empty for $\mu(A) > 0$ even if T is not ergodic. These are also known in the literature as *intersective* sets. (See [Ruz82]). Much is known about these (see [Fur81], [B-M86], [BH96], [BFM96]). In particular $\{n^r; n = 1, 2, \ldots\}$ is a Poincaré set for each $r = 1, 2, 3, \ldots$

For a family $\mathcal{F}$ of subsets of $\mathbb{Z}$ it is customary to denote by $\mathcal{F}^*$ the dual family: $\mathcal{F}^* = \{S \subset \mathbb{Z} : \forall S' \in \mathcal{F}, S \cap S' \neq \emptyset\}$. Note that $\{\text{syndetic}\} = \{\text{thick}\}^*$, $\{\text{thick}\} = \{\text{syndetic}\}^*$ and $\{RT\} = \{\text{Poincaré}\}^*, \{\text{Poincaré}\} = \{\text{RT}\}^*$.

We have seen above that a Δ_r^*-set is necessarily syndetic. One of our objectives is to sharpen this statement.

We will need the notion of a "PW-$\mathcal{F}$" set for a family $\mathcal{F}$ of subsets of $\mathbb{Z}$. "PW" stands for "piecewise" and if $S \in \mathcal{F}$ and Q is a thick set then we shall say $S \cap Q$ is PW-$\mathcal{F}$ (or $S \cap Q \in$ PW-$\mathcal{F}$). Clearly this notion is useful only for families of syndetic sets. "PW-syndetic" is itself a useful notion. Van der Waerden's theorem [GRS80] implies that syndetic sets contain arbitrarily long arithmetic progressions. In fact this is true for PW-syndetic sets. Unlike the family of syndetic sets, the latter have the "divisibility" property: if S is PW-syndetic and $S = S_1 \cup S_2 \cup \cdots \cup S_k$ is a finite partition, then some S_i is PW-syndetic, see [Bro71]. A recent result of Renling Jin [Jin02] is the following:

Theorem 1.2. *If $A, B \subset \mathbb{Z}$ and $d^*(A) > 0$, $d^*(B) > 0$, then $A + B$ is PW-syndetic.*

We will sharpen this to

Theorem I. *If $A, B \subset \mathbb{Z}$ and $d^*(A) > 0$, $d^*(B) > 0$, then $A + B$ is a PW-Bohr set (PWB-set).*

In particular $d^*(A) > 0$ will imply that $A - A$ is a PW-Bohr set. More precisely it is a PW-Bohr$_0$ (PWB$_0$)-set. This will also follow from our earlier observation that it is a Δ_r^*-set for sufficiently large r, and from

Theorem II. *For each $r \geq 2$, a Δ_r^*-set is PW-Bohr$_0$.*

It is not hard to see that the prefix "PW" is indispensable in these theorems. For example $A = \bigcup[10^n, 10^n + n]$ has $d^*(A) = 1$ but $A + A$ is not syndetic. Also since $x^3 + y^3 = z^3$ has no solution in non-zero integers, it follows that the set of non-cubes $S = \mathbb{Z}\backslash\{n^3; n = \pm1, \pm2, \pm3, \ldots\}$ is a Δ_3^* set. But by Weyl's equidistribution theorem S is not a Bohr_0-set. (See Theorem 4.1 below for a stronger form of this observation.)

From Theorem I we shall deduce the following result which should be compared with a theorem due to Ruzsa ([Ruz82], Theorem 3) which states that if $d^*(A) > 0$, then $A + A - A$ is a Bohr set. (Both Ruzsa's theorem and our result can be viewed as improvements on a theorem of Bogoliouboff ([Bog39], [Føl54]) which implies that if $d^*(A) > 0$, then $A - A + A - A$ is a Bohr set.)

Corollary 1.3. *If A, B, C are three subsets of $\mathbb{Z}$ with positive upper density and one of them is syndetic, then $A + B + C$ is a Bohr set.*

2 Measure Preserving Systems, Time Series, and Generic Schemes

In this section we introduce a basic tool which will be needed repeatedly: the correspondence between data given on large intervals of time ("time series") and measure preserving dynamical systems. This tool has been used previously under the name "correspondence principle" (see e.g., [Ber96]) and here we present it in a more general form. We repeat the definition of a measure preserving system which was given informally in §1.

Definition 2.1. *A* measure preserving system *is a quadruple $(X, \mathcal{B}, \mu, T)$ where $(X, \mathcal{B}, \mu)$ is a probability space where we assume $\mathcal{B}$ is countably generated, and T is a measurable, invertible, and measure preserving map, $T\colon X \to X$. The system is* ergodic *if every measurable T-invariant set has measure* 0 *or* 1.

For a measurable function $f\colon X \to \mathbb{C}$ we denote by Tf the function $Tf(x) = f(Tx)$. We take note of the ergodic theorem (see, for example, [Kre85]):

Theorem 2.2. *If $(X, \mathcal{B}, \mu, T)$ is a measure preserving system and $f \in L^1(X, \mathcal{B}, \mu)$, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} T^n f = \overline{f}$$

exists almost everywhere. If $f \in L^p(X, \mathcal{B}, \mu)$, $1 \le p < \infty$, the convergence is in L^p as well. If the system is ergodic then $\overline{f} = \int f d\mu$ a.e., so that the average of the sequence $\{f(T^n x)\}$ equals a.e. the average of f over X.

Sequences of the form $\{f(T^n x)\}_{a \le n \le b}$ are referred to as "time series". In a certain sense the ergodic theorem enables one to reconstruct a dynamical system from "time series data". We shall make this precise in the notion of "generic schemes" which we proceed to define. In the next definitions the indices l and r range over the natural numbers.

Definition 2.3. *An* array *is a sequence* $\{J_l\}$ *of intervals of integers,* $J_l = \{a_l, a_l+1, \ldots, b_l\}$ *for which* $|J_l| = b_l - a_l + 1 \to \infty$ *as* $l \to \infty$.

Definition 2.4. *A* scheme $(\{J_l\}, \{\xi_r^l\})$ *is an array* $\{J_l\}$ *together with a doubly indexed set of complex-valued functions* $\{\xi_r^l\}$ *where, for each* r, $\xi_r^l(n)$ *is defined for* $n \in J_l$ *and, for each* r, *the functions* $\{\xi_r^l; l = 1, 2, \ldots\}$ *are uniformly bounded. For* $n \notin J_l$ *we take* $\xi_r^l(n) = 0$. *The* $\{\xi_r^l\}$ *will be referred to as* time series. *They are defined on all of* $\mathbb{Z}$ *but only the values on* J_l *have significance. The following notion relates closely to that of a "stationary stochastic process".*

Definition 2.5. *A* process $(X, \mathcal{B}, \mu, T, \Phi)$ *consists of a measure preserving system* $(X, \mathcal{B}, \mu, T)$ *together with an at most countable ordered set* $\Phi = \{\varphi_1, \varphi_2, \ldots\}$ *of* L^∞*-functions on* X *such that* $\mathcal{B}$ *is the* σ*-algebra generated by the functions of* Φ *and their translates under* T. *(When the* φ_i *are complex valued we assume* Φ *closed under conjugation). A process is ergodic if the underlying measure preserving system is ergodic.*

Finally we have

Definition 2.6. *A scheme* $(\{J_l\}, \{\xi_r^l\})$ *is* generic *for a process* $(X, \mathcal{B}, \mu, T, \Phi)$ *if for every* m *and for every choice of* $i_1, i_2, \ldots, i_m$ *and* $j_1, j_2, \ldots, j_m$ *(the indices here need not be distinct):*

$$\lim_{l \to \infty} \frac{1}{|J_l|} \sum_{n \in J_l} \xi_{i_1}^l(n+j_1)\xi_{i_2}^l(n+j_2)\cdots\xi_{i_m}^l(n+j_m) \tag{1}$$
$$= \int_X T^{j_1}\varphi_{i_1} T^{j_2}\varphi_{i_2} \cdots T^{j_m}\varphi_{i_m}\, d\mu$$

It will be convenient to introduce the countable family Φ^* consisting of the products appearing in (1):

$$\Phi^* = \{\psi = T^{j_1}\varphi_{i_1} T^{j_2}\varphi_{i_2} \cdots T^{j_m}\varphi_{i_m}\}$$

The corresponding time series have the form

$$\zeta^l(n) = \xi_{i_1}^l(n+j_1)\xi_{i_2}^l(n+j_2)\cdots\xi_{i_m}^l(n+j_m),$$

and when (1) holds, we say that $\{\zeta^l\}$ *represents* ψ.

It will be convenient in the sequel to regard Φ^* as the increasing union of finite sets, $\Phi^* = \bigcup_{h=1}^{\infty} \Phi_h^*$. The subscript h has no significance other than as an index with $\Phi_1^* \subset \Phi_2^* \subset \cdots \subset \Phi_h^* \subset \cdots$.

We note that the ergodic theorem implies that if $(X, \mathcal{B}, \mu, T)$ is ergodic, then for almost every $x_0 \in X$, the scheme $(\{J_l\}, \{\xi_r^l\}$ is generic for the process $(X, \mathcal{B}, \mu, T, \Phi)$ with $J_l = [1, l]$ and $\xi_r^l(n) = \varphi_r(T^n x_0)$ independently of l.

The main result of this section goes in the opposite direction, and will attach to an arbitrary scheme an ergodic process. First we need the notions of *subarrays* and *subschemes.*

Definition 2.7. *An array $\{H_l\}$ is a* subarray *of $\{J_l\}$ if $l \to L_l$ is a monotone increasing function from $\mathbb{N}$ to $\mathbb{N}$ and H_l is a subinterval of J_{L_l}.*

Definition 2.8. *A scheme $(\{H_l\}, \{\eta_r^l\})$ is a* subscheme *of $(\{J_l\}, \{\xi_r^l\})$ if $\{H_l\}$ is a subarray of $\{J_l\} : H_l \subset J_{L_l}$, and η_r^l is the restriction of $\xi_r^{L_l}$ to H_l.*

Our main result in this section is

Theorem 2.9. *For any scheme $(\{J_l\}, \{\xi_r^l\})$ there exists a subscheme and an ergodic process for which the subscheme is generic.*

Proof. First we will pass to a subscheme which is generic for a process $(X, \mathcal{B}, \mu, T, \Phi)$ which is not necessarily ergodic. For each r, let $\Lambda_r \subset \mathbb{C}$ be a compact set with $\xi_r^l(n) \in \Lambda_r$ for all l and n. Let $\tilde{\Lambda} = \prod \Lambda_r$ and let $X = \tilde{\Lambda}^{\mathbb{Z}}$. We denote by ξ_r^l the point in $\Lambda_r^{\mathbb{Z}}$ with $\xi_r^l = (\ldots, \xi_r^l(-1), \xi_r^l(0), \xi_r^l(1), \ldots)$ and form $\tilde{\xi}^l = (\xi_1^l, \xi_2^l, \ldots) \in \tilde{\Lambda}^{\mathbb{Z}} = X$. X is a compact metrizable space and we form the measures

$$\nu_l = \frac{1}{|J_l|} \sum_{n \in J_l} \delta_{T^n \tilde{\xi}^l} \tag{2}$$

where $T: X \to X$ denotes the shift map $T\omega(n) = \omega(n+1)$. Since $|J_l| \to \infty$, any weak limit of a subsequence of ν_l is T-invariant, and we let ν be some such limit: $\nu = \lim \nu_{L_l}$. It is not hard to see that $(\{J_{L_l}\}, \{\xi_r^{L_l}\})$ is generic for the process $(X, \mathcal{B}, \nu, T, \Phi)$ where $\mathcal{B}$ is the Borel σ-algebra of sets in X and $\Phi = \{\varphi_1, \varphi_2, \ldots\}$ with φ_r the functions on $\tilde{\Lambda}^{\mathbb{Z}}$ given by $\varphi_r(\omega) = \omega(0)(r)$. By ergodic decomposition there will be an ergodic measure μ whose support is a subset of the support of ν. Any point in the support of μ is a limit of points of the form $T^n \tilde{\xi}^l$ with $n \in J_l$ and $l \to \infty$, by (2). Since μ is ergodic, almost every point ω in its support is *generic* for μ, in the sense that averages of a given bounded measurable function along the orbit of ω tend to the integral of the function. In particular for functions in Φ^* we have:

$$\frac{1}{N} \sum_{n=k}^{k+N-1} T^{j_i} \varphi_{i_1} T^{j_2} \varphi_{i_2} \cdots T^{j_m} \varphi_{i_m}(T^n \omega) \longrightarrow \int T^{j_1} \varphi_{i_1} T^{j_2} \varphi_{i_2} \cdots T^{j_m} \varphi_{i_m} d\mu \tag{3}$$

uniformly for $|k| \leq N$.

We can find N sufficiently large that the difference of the two sides in (3) is $< \varepsilon$ for all $T^{j_1} \varphi_{i_1} \cdots T^{j_m} \varphi_{i_m} \in \Phi_h^*$. We then choose $T^n \tilde{\xi}^l$ close enough to ω, $n \in J_l$, so that the difference of the two sides of (3) remains $< \varepsilon$ with ω replaced by $T^n \tilde{\xi}^l$. Since $n \in J_l$, assuming l sufficiently large, we will have

$H_l = [n+k, n+k+N-1] \subset J_l$ for some k with $|k| \leq N$. We now let $\varepsilon \to 0$, $h \nearrow \infty$, and choose an appropriate subsequence of l; rescrambling the information in (3) we find a subscheme $(\{H_l\}, \{\xi_r^l\})$ which is generic for $(X, \mathcal{B}, \mu, T, \Phi)$. ◻

Scholium to Theorem 2.9. If for some r,

$$\limsup_{l\to\infty} \frac{1}{|J_l|} \Big| \sum_{n\in J_l} \xi_r^l(n) \Big| > 0,$$

we can add the condition that the corresponding φ_r does not vanish a.e. This follows from the fact that the measure ν satisfies $\int \varphi_r d\nu \neq 0$ and so ν must have an ergodic component with $\int \varphi_r d\mu \neq 0$.

We remark that in the case of ergodic processes, given a generic scheme, "many" subschemes will again be generic. This is made precise in the following: For any process $(X, \mathcal{B}, \mu, T, \Phi)$, Φ^* is countable and we fix an increasing family of finite sets $\Phi_h^* \subset \Phi^*$ increasing to Φ^*. Given a scheme $(\{J_l\}, \{\xi_r^l\})$ and fixing l, and letting $\varepsilon > 0$, we shall say that an interval $H \subset J_l$ is *ε-h-generic* for the process $(X, \mathcal{B}, \mu, T, \Phi)$ if (1) holds approximately; i.e, if for every $\psi \in \Phi_h^*$ and corresponding time series $\zeta^l(n)$.

$$\left| \frac{1}{|H|} \sum_{n\in H} \zeta^l(n) - \int \psi d\mu \right| < \varepsilon. \tag{4}$$

Assume now a process $(X, \mathcal{B}, \mu, T, \Phi)$ given with $\Phi^* = \bigcup \Phi_h^*$ as above, and let $(\{T_l\}, \{\xi_r^l\})$ be a generic scheme for the process.

Proposition 2.10. *If $(X, \mathcal{B}, \mu, T, \Phi)$ is an ergodic process, then for any $\varepsilon > 0$ and $h \in \mathbb{N}$ there exists $p_0 \in \mathbb{N}$ so that for any $p \geq p_0$ there exists a positive number $l_0(\varepsilon, h, p)$ so that for $l > l_0(\varepsilon, h, p)$, at least $(1-\varepsilon)(|J_l| - p + 1)$ of the $(|J_l| - p + 1)$ intervals of length p in J_l are ε-h-generic for the process.*

Letting p and l grow we see, according to the proposition, that the intervals J_l can be replaced by many choices of subintervals, and the scheme will remain generic. It is easy to see that this is not true for non-ergodic processes (where time series have different statistical behavior along different intervals of time).

Proof of Proposition 2.10. It suffices to treat a single function and the corresponding time series. For if for each of the $|\Phi_h^*|$ functions in Φ_h^* we have $(1-\varepsilon_1)(|J_l| - p + 1)$ "ε_1-generic" intervals with $\varepsilon_1 |\Phi_h^*| < \varepsilon$, the number of intervals common to all of these will not be less than $(1-\varepsilon)(|J_l| - p + 1)$, and these intervals are ε_1-h-generic, and so also ε-h-generic. So let $\psi \in \Phi^*$. Ergodicity assures that for p large, $\frac{1}{p} \sum_{q=0}^{p-1} T^q \psi$ is L^2-close to $\int \psi d\mu$, and so

$$\int \left(\frac{1}{p} \sum_{q=0}^{p-1} T^q \psi \right)^2 d\mu - \left(\int \psi d\mu \right)^2$$

is small. Fix p and set $\eta(n) = \frac{1}{p}\sum_{q=0}^{p-1} \zeta(n+q)$. η and ζ have the same long-term averages,

$$\frac{1}{|J_l|}\sum_{n\in J_l}\left(\eta(n) - \int \psi d\mu\right)^2 = \frac{1}{J_l}\sum_{n\in J_l}\eta(n)^2 - 2\left(\frac{1}{|J_l|}\sum_{n\in J_l}\eta(n)\right)\left(\int \psi d\mu\right) + \left(\int \psi d\mu\right)^2$$

$$\longrightarrow \int\left(\frac{1}{p}\sum_{q=0}^{p-1} T^q\psi\right)^2 d\mu - \left(\int \psi d\mu\right)^2$$

which is small for large p. But this implies that most $\eta(n)$ are close to $\int \psi d\mu$ as asserted in the proposition. □

3 Some Examples of PW-Bohr Sets

3.1 Fourier Transforms

Our first example of PW-Bohr sets will lead to three more in the following subsections.

Theorem 3.1. *Let ω be a non-negative measure on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ with a non-trivial discrete (atomic) component, and let $\hat{\omega}$ denote its Fourier transform: $\hat{\omega}(n) = \int_{\mathbb{T}} e^{2\pi itn} d\omega(t)$. If*

$$S = \{n : \mathrm{Re}\,\hat{\omega}(n) > 0\},$$

then S is a PW-Bohr$_0$ set.

Proof. Let ω_d denote the discrete component of ω: $\omega_d = \sum_{\lambda\in\Lambda} \omega(\{\lambda\})\delta_\lambda$ where Λ consists of all the atoms of ω . Let Λ_0 be a finite subset of Λ so that $\omega_d(\Lambda_0) > \frac{3}{4}\omega_d(\Lambda)$. Set

$$\psi(\tau) = \sum_{\lambda\in\Lambda_0} \omega_d(\lambda)e^{2\pi i\lambda\tau}$$

and let B_0 be the Bohr$_0$ set: $B_0 = \{n : \mathrm{Re}\,\psi(n) > \frac{2}{3}\omega_d(\Lambda_0)\}$. The measure $\omega - \omega_d$ is continuous and so by Wiener's theorem (see [Kre85], p.96)

$$\lim_{N\to\infty} \frac{1}{2N+1}\sum_{n=-N}^{N}\left|\hat{\omega}(n) - \hat{\omega}_d(n)\right|^2 = 0$$

It follows that $Q' = \left\{n : \left|\hat{\omega}(n) - \hat{\omega}_d(n)\right| \geq \frac{1}{3}\omega_d(\Lambda_0)\right\}$ has density 0 so that $Q = \mathbb{Z}\backslash Q'$ is a thick set.

In $B_0 \cap Q$,

$$\begin{aligned}\operatorname{Re}\hat{\omega}(n) &> \operatorname{Re}\hat{\omega}_d(n) - \tfrac{1}{3}\omega_d(\Lambda_0)\\ &\geq \operatorname{Re}\psi(n) - \omega_d(\Lambda\backslash\Lambda_0) - \tfrac{1}{3}\omega_d(\Lambda_0)\\ &> \operatorname{Re}\psi(n) - \tfrac{1}{4}\omega_d(\Lambda) - \tfrac{1}{3}\omega_d(\Lambda_0)\\ &> \tfrac{2}{3}\omega_d(\Lambda_0) - \tfrac{1}{3}\omega_d(\Lambda_0) - \tfrac{1}{3}\omega_d(\Lambda_0) = 0\end{aligned}$$

so that $B_0 \cap Q \subset S$. It follows that S is PWB_0. □

3.2 Positive Definite Sequences

Theorem 3.2. *Let $\{a(n)\}_{n\in\mathbb{Z}}$ be a positive definite sequence of non-negative reals for which* $\lim \frac{1}{2N+1} \sum_{n=-N}^{N} a(n) > 0$. *Then $S = \{n : a(n) > 0\}$ is a PWB_0 set.*

Proof. By Herglotz's theorem $a(n) = \hat{\omega}(n)$ for some non-negative measure ω on $\mathbb{T}$ [Hel83], and the hypothesis of the theorem implies that $\omega\{0\} > 0$. The previous theorem applies and so S is PWB_0. □

3.3 Return Time Sets

A consequence of the foregoing is that RT-sets are PW-Bohr_0 sets. Recall a return time set has the form $S \supset R(A) = \{n : \mu(A \cap T^{-n}A) > 0\}$ where $(X, \mathcal{B}, \mu, T)$ is a measure preserving system, $A \in \mathcal{B}$ and $\mu(A) > 0$. If $a(n) = \mu(A \cap T^{-n}A)$ we can write $a(n) = \int fT^n f d\mu$ with $f = 1_A$ and T is a unitary operator. It is easily checked that $\sum_{m,n=1}^{N} a(n-m)x_n\overline{x}_m \geq 0$ for any $x_1, x_2, \ldots, x_N$, and so $\{a(n)\}$ is a positive definite sequence. We also have

$$\frac{1}{2N+1}\sum_{n=-N}^{N}\int fT^n f d\mu \longrightarrow \int fP_T f d\mu,$$

where P_T is the self-adjoint projection of $L^2(X, \mathcal{B}, \mu)$ to the subspace of T-invariant functions. Since $\int P_T f d\mu = \mu(A)$, it follows that $P_T f \neq 0$, and since $\int fP_T f d\mu = \int fP_T^2 f d\mu = \int (P_T f)^2 d\mu > 0$ the hypotheses of Theorem 3.2 are fulfilled. This proves

Theorem 3.3. *RT sets are PW-$Bohr_0$.*

3.4 Difference Sets of Sets of Positive Upper Density

Proposition 3.4. *Let $\{J_l\}$ be an array and, for each l, let $S_l \subset J_l$ with $|S_l| > \delta|J_l|$ for fixed $\delta > 0$. Then $\bigcup(S_l - S_l)$ is PW-$Bohr_0$.*

This leads immediately to

Theorem 3.5. *If $d^*(S) > 0$, then $S - S$ is PWB_0 for $S \subset \mathbb{Z}$*

Proof of Proposition 3.4. We form the scheme $(\{J_l\}, \{\xi^l\})$, where the usual index r is suppressed since it takes only one value, and we define $\xi^l(n) = 1_{S_l}(n)$. We pass to a subscheme which is generic for a process $(X, \mathcal{B}, \mu, T, \{\varphi\})$ where, according to the scholium following Theorem 2.2, φ is not almost everywhere 0. By the construction ($\Lambda = \{0, 1\}$), φ takes on the values 0, 1 and so $\varphi = 1_A$ for $A \in \mathcal{B}$, $\mu(A) > 0$. By definition of a generic scheme

$$\mu(A \cap T^{-k}A) = \int \varphi T^k \varphi d\mu = \lim \frac{1}{|H_l|} \sum_{n \in H_l} \xi^l(n)\xi^l(n+k)$$

which will be > 0 only if $k \in \bigcup(S_l - S_l)$. This proves the proposition. ◻

In the sequel we will use a stronger version of Proposition 3.4. Let us say that a set Q is *uniformly thick* if for every $l \in \mathbb{N}$, $\exists l' \in \mathbb{N}$ so that every interval J of length l' meets Q in a set containing an interval of length l. This will happen if $\frac{1}{N} \sum_{j=n+1}^{n+N} 1_Q(j) \to 1$ uniformly in n. If ω is a continuous measure on $\mathbb{T}$ then Wiener's Theorem can be sharpened to

$$\frac{1}{N} \sum_{j=n+1}^{n+N} |\hat{\omega}(j)|^2 \longrightarrow 0$$

uniformly in n. Using this in the proof of Theorem 3.1 we find that the set S of that theorem is the intersection of a Bohr_0-set and a uniformly thick set. If we call a set a UPW-Bohr_0 set if it contains intersection of a Bohr set and a uniformly thick set, we can replace PW-Bohr_0 throughout this section by UPW-Bohr_0. For later reference we re-write Proposition 3.4 in its strengthened form as

Proposition 3.6. *Let $\{J_l\}$ be an array and for each l, let $S_l \subset J_l$ with $|S_l| > \delta|J_l|$ for fixed $\delta > 0$. Then $\bigcup(S_l - S_l)$ is a UPW-$Bohr_0$ set.*

4 The Hierarchy of Families of Large Sets

We consider the following families of "large sets":

(a) B_0 = Bohr_0 sets
(b) RT = return time sets
(c) $\bigcup \Delta_r^*$ = sets which for some r meet every $(S - S)\backslash\{0\}$ provided $|S| \geq r$
(d) PWB_0 = piecewise Bohr_0 sets
(e) PWB = piecewise Bohr sets
(f) PW Syn = piecewise syndetic sets
(g) PD = sets of positive upper Banach density = $\{S : d^*(S) > 0\}$

It is easily seen that the first of these families is contained in the second, the second in the third, the fourth in the fifth and the fifth in the sixth. That $\bigcup \Delta_r^* \subset \mathrm{PWB}_0$ is the content of our Theorem II to be proved in §9. In fact all these inclusions are proper, and in this section we shall show that (b) $\neq$ (c), (c) $\neq$ (d) and (e) $\neq$ (f). The fact that (a) $\neq$ (b) follows from work of I. Kříž [Kříž87] and that (f) $\neq$ (g) is an exercise.

Theorem 4.1. *There are Δ_3^*-sets which do not contain RT-sets. So (b) $\neq$ (c).*

Proof. We use the fact ([Fur81], [Sár78]) that for every $r = 1, 2, \ldots$ the set $P_r = \{n^r\}_{n\in\mathbb{Z}}$ is a Poincaré set; i.e., it meets every return time set. Hence $\mathbb{Z}\backslash P_r$ does not contain any RT-set. On the other hand, when $r \geq 3, \mathbb{Z}\backslash P_r$ is a Δ_3^*-set. For, by Fermat's theorem, for any distinct a, b, c, we cannot have $b - a$, and $c - b$ as well as $c - a = (b - a) + (c - b)$ all in P_r. $\square$

To prove that (c) $\neq$ (d) we produce a set of density 0 in $\mathbb{Z}$ that contains a Δ_r-set for every r. The complement of this set cannot belong to any Δ_r^*. On the other hand, the complement of a set of density 0 contains arbitrarily long intervals, and so is thick, and in particular it is PWB_0. So we take as a Δ_r-set a set of the form

$$D_r = \{-rq_r, -(r-1)q_r, \ldots, -q_r, 0, q_r, \ldots, (r-1)q_r, rq_r\}$$

Choosing $q_r = r^3$ we can check that the density of $\bigcup D_r$ is 0. This proves

Theorem 4.2.
$$\bigcup \Delta_r^* \neq PWB_0$$

Finally we have (e) $\neq$ (f) by the following:

Theorem 4.3. *There are syndetic sets that are not PWB.*

Proof. We use considerations from topological dynamics. Let $\Omega = \{0,1\}^{\mathbb{Z}}$ and define the shift T on Ω by $T\omega(n) = \omega(n+1)$. If $M \subset \Omega$ is a minimal closed T-invariant subset, $M \neq \{0\}$, then for any $\omega \in M$, $\{n : \omega(n) = 1\}$ is syndetic. We can choose M so that the system (M, T) is weakly mixing ([Fur81]). Let $\xi \in M$ and set $S = \{n : \xi(n) = 1\}$. Assume S is PWB; then $S = Q \cap P$ where Q is thick and P is a Bohr set. If $\eta = 1_P$ then ξ and η agree on arbitrarily long intervals and for some $\{n_k\}, \lim T^{n_k}\xi = \lim T^{n_k}\eta$. Let $L = \overline{\{T^n\eta\}_{n\in\mathbb{Z}}}$ be the closed invariant set generated by η (so that $M \cap L \neq \emptyset$). Since M is minimal, $M \subset L$. By definition of a Bohr set there is a torus $\mathbb{T}^m$, a rotation $R : \mathbb{T}^m \to \mathbb{T}^m$, $R(\theta) = \theta + \alpha$, and an open set $U \subset \mathbb{T}^m$ so that $R^n(0) \in U \Rightarrow \eta(n) = 1$. Let $A = \{\omega : \omega(0) = 1\}$; then $R^n(0) \in U \Rightarrow T^n\eta \in A$. Let $Z \subset \mathbb{T}^m$ be the closed subgroup of $\mathbb{T}^m$ generated by α. By [Fur67] (Z, R) and (M, T) are disjoint, and since both are minimal, $Z \times M$ is minimal for $R \times T$. This implies that $\{(R^n(0), T^n\eta)\}$ is dense in $Z \times M$. But from the foregoing, when the first coordinate is in U the other is in A. It follows that $U \times A$ is dense in $U \times M$; hence $M = A$ and $\xi \equiv 1$. Choosing M non-degenerate gives us the example we seek. $\square$

5 The Sum Set of Positive Density Sets

In this section we will prove Theorem I which asserts that the sum set of two sets A, B with positive upper density is a PW-Bohr set.

We begin with an elementary lemma.

Lemma 5.1. *Let $J, J' \subset \mathbb{Z}$ be intervals of length l, l' respectively. Let $S \subset J$, $S' \subset J'$ be subsets satisfying $|S| \geq \delta l$, $|S'| \geq \delta' l'$. We can find an interval L and a subset $R \subset L$ so that for some t, $S + S' \supset R - R + t$ and such that $|R| \geq \frac{\delta\delta'}{2}|L|$.*

Proof. Without loss of generality we suppose $l \leq l'$. For each $t \in \mathbb{Z}$, form $R_t = S \cap (t - S')$. $|R_t|$ equals the number of points of $S \times S'$ lying on the line $x + y = t$. The number of such lines meeting $S \times S'$ doesn't exceed $l + l'$, and so for some t,

$$|R_t| \geq \frac{|S \times S'|}{l + l'} \geq \frac{\delta\delta' l l'}{l + l'} \geq \frac{\delta\delta'}{2} l.$$

Take $R = R_t$ so that $R - R \subset S + (S' - t)$, and take $L = J$. □

Theorem I will now follow from

Theorem 5.2. *Let $\{J_k\}$ be an array (Def. 2.3), and let $S_k \subset J_k$, with $|S_k| > \delta|J_k|$ where $\delta > 0$. Let $\{t_k\}$ be an arbitrary set of integers. The set $A = \bigcup_{k=1}^{\infty} (S_k - S_k + t_k)$ is PW-Bohr.*

Our next step is to reduce Theorem 5.2 to a special case in which the sets S_k are related. For two sets of integers S', S'', let us write $S' \prec S''$ if for some $c \in Z$, $S' + c \subset S''$. Clearly $S' \prec S''$ implies that $S' - S' \subset S'' - S''$.

Lemma 5.3. *Theorem 5.2 is true in general if it is true for the case that $S_k \prec S_{k+1}$ for each $k = 1, 2, 3, \ldots$.*

Proof. We consider the general case of an arbitrary array $\{J_k\}$ with subsets $S_k \subset J_k$. We follow the procedure in the proof of Proposition 3.4 based on Theorem 2.2 to obtain a subscheme of $(\{J_k\}, \{1_{S_k}\})$ generic for an ergodic process $(X, \mathcal{B}, \mu, T, 1_A)$ with $\mu(A) > 0$. Reindexing and renaming sets we suppose that $(\{J_k\}, \{1_{S_k}\})$ is generic for the above process. Note that the hypothesis of genericity implies that we will still have $|S_k| > \delta'|J_k|$ for some positive δ'. We now pass to a further subscheme for which $S'_k \prec S'_{k+1}$. This is done as follows. Removing a set of measure 0 from A we can assume that any non-empty intersection $A \cap T^{-\tau_1}A \cap T^{-\tau_2}A \cap \cdots \cap T^{-\tau_r}A$ has positive measure. It follows from the ergodic theorem that there exist points x with $T^{\tau_n}x \in A$ for a sequence $\tau_1 < \tau_2 < \cdots < \tau_n < \cdots$ (depending on x) with $\lim \frac{\tau_n}{n} < \infty$. Thus $A \cap T^{-\tau_1}A \cap T^{-\tau_2}A \cap \cdots \cap T^{-\tau_r}A$ is non-empty for each r and by our assumption $\mu(A \cap T^{-\tau_1}A \cap T^{-\tau_2}A \cap \cdots \cap T^{-\tau_r}A) > 0$ for each r. By genericity of $(\{J_k\}, \{1_{S_k}\})$ this implies that translating $\{0, \tau_1, \tau_2, \cdots, \tau_r\}$ by some c_r we will obtain a subset of some $S_k : \{c_r, c_r+\tau_1, c_r+\tau_2, \cdots, c_r+\tau_r\} \subset S_{k(r)}$. We now set $J'_r = [c_r, c_r+\tau_r] \subset J_{k(r)}$ and $S'_r = \{c_r, c_1+\tau_1, c_r+\tau_2, \cdots, c_r+\tau_r\}$. Then $S'_r \prec S'_{r+1}$ and since $S'_r \subset S_{k(r)}$, $\bigcup(S_k - S_k + t_k) \supset \bigcup(S'_r - S'_r + t_{k(r)})$. At the same time $\lim \frac{\tau_n}{n} < \infty$ so that $\exists \alpha > 0$ with $r = |S'_r| \geq \alpha |J'_r|$. ☐

We show now how Theorem 5.2 follows from Proposition 3.6.

Proof of Theorem 5.2. According to the foregoing lemma, we may assume that for each k, $S_k - S_k \subset S_{k+1} - S_{k+1}$. For each $m = 1, 2, 3, \ldots$, let $k(m)$ be chosen so that $(S_k - S_k) \cap [-m, m]$ is a fixed set for $k \geq k(m)$. Write $S'_m = S_{k(m)}$ and $t'_m = t_{k(m)}$; we will show that $\bigcup(S'_m - S'_m + t'_m)$ is PW-Bohr. By Proposition 3.6 $\bigcup(S'_m - S'_m)$ is UPW-Bohr$_0$; i.e., it contains the intersection of a Bohr$_0$ set H and a *uniformly* thick set Q. Thus there is a trigonometric polynomial $\psi(t) = \sum_{j=1}^{N} a_j e^{i\lambda_j t}$ with $\operatorname{Re}\psi(0) > 0$ such that for any $n \in Q$, if $\operatorname{Re}\psi(n) > 0$ then $n \in \bigcup(S'_m - S'_m)$. Form $\psi_m(t) = \psi(t - t'_m)$ and pass to a subsequence $\{m_p\}$ so that these converge uniformly to a polynomial $\psi'(t)$. Let $0 < \alpha < \operatorname{Re}\psi(0)$. By almost periodicity of $\psi(t)$ it follows that $\operatorname{Re}\psi'(n) > \alpha$ on a non-empty (and therefore syndetic) set of n. We can suppose that the subsequence $\{m_p\}$ is such that $\operatorname{Re}\psi'(n) > \alpha$ implies $\operatorname{Re}\psi(n - t'_{m_p}) > 0$ for each p. Form the set $Q' = \bigcup([-m_p, m_p] \cap Q + t'_{m_p})$. Suppose $\operatorname{Re}\psi'(n) > \alpha$ with $n \in Q'$. Then for some p, $n - t'_{m_p} \in [-m_p, m_p] \cap Q$ and $\operatorname{Re}\psi(n - t'_{m_p}) > 0$. It follows that $n - t'_{m_p} \in (\bigcup(S'_m - S'_m)) \cap [-m_p, m_p]$. By the choice of $\{S'_m\}$ this implies $n \in S'_{m_p} - S'_{m_p} + t'_{m_p}$. Since Q is uniformly thick, for large p, $[-m_p, m_p] \cap Q$ contains large intervals and this implies that Q' is a thick set. This proves that $\bigcup(S'_m - S'_m + t'_m)$ is a PW-Bohr set. ☐

This completes the proof of Theorem I.

Corollary 5.4 (Corollary 1.3 of §1). *If $A, B, C, \subset \mathbb{Z}$ are three sets with positive upper density, one of which is syndetic, then $A+B+C$ is a Bohr set.*

This will follow from the Theorem I together with the following lemma:

Lemma 5.5. *If R is a PW-Bohr set and S is syndetic in $\mathbb{Z}$ then $R+S$ is Bohr.*

Proof. A translate of R will be PW-Bohr$_0$ and the opposite translate of S is syndetic, so we can assume that R is a PW-Bohr$_0$ set. This means that there is a torus $\mathbb{T}^m$, an $\alpha \in \mathbb{T}^m$, a neighborhood U of 0 in $\mathbb{T}^m$ and a thick set Q with $R \supset \{n : n\alpha \in U\} \cap Q$. Let V be a neighborhood of 0 in $\mathbb{T}^m$ with $V - V \subset U$ and let $\beta_1, \beta_2, \ldots \beta_k \in \mathbb{T}^m$ so that $\mathbb{T}^m = \bigcup_{l=1}^{k}(\beta_l + V)$.

We claim that for some $l, 1 \leq l \leq k$, $S + R \supset \{n : n\alpha \in \beta_l + V\}$ which implies that $S+R$ is a Bohr set. Assume this isn't so; then for each l, $\exists x_l$ with $x_l\alpha \in \beta_l + V$ and $x_l \notin S+R$. Let $S_l = S \cap \{n : n\alpha \in \beta_l + V\}$ so that $S = \bigcup S_l$. We have $x_l \notin S+R$ and so $x_l - S_l \cap R = \varnothing$. Since $x_l\alpha \in \beta_l + V$ and $S_l\alpha \subset \beta_l + V$ we have $(x_l - S_l)\alpha \subset U$. Now $R \supset \{n : n\alpha \in U\} \cap Q$ so $(x_l - S_l) \cap R = \varnothing$

implies that $(x_l - S_l) \subset Q^c$, the complement of Q. Equivalently $S_l \subset (x_l - Q)^c$, so $S = \bigcup S_l \subset \left(\bigcap (x_l - Q)\right)^c$. But the intersection of finitely many translates of a thick set is thick whereas S is syndetic. This contradiction proves our assertion. □

6 Kronecker-complete Processes

The remaining sections are directed to giving a proof of Theorem II of §1. The crucial step in this proof is a proposition to be proved in §8 which generalizes the fact (Theorem 3.5) that $d^*(A) > 0$ implies that $A - A$ is PWB_0. To achieve this generalization we will use once more the correspondence described in §2 between schemes and processes. Another ingredient that will enter is the point spectrum of an ergodic system, i.e., the eigenvalues of the operator T on the L^2-space of the system. It will be of importance that in a scheme generic for a process for which non-trivial eigenvalues exist, the eigenfunctions are also represented. This leads to the notion dealt with in this section of a "Kronecker-complete process."

We begin by recalling the notion of the "Kronecker factor" of an ergodic system: Let $(X, \mathcal{B}, \mu, T)$ be an ergodic measure preserving system. There is a compact abelian group Z and an element $\alpha \in Z$ whose multiples $\{n\alpha\}$ are dense in Z, and a map $\pi\colon X \to Z$ which is measurable and measure preserving with respect to Haar measure dz on Z, and such that for a.e. $x \in X$, $\pi(Tx) = \pi(x) + \alpha$. If $\chi \in \hat{Z}$ is a character on Z then $f = \chi \circ \pi$ is an eigenfunction of T: $f(Tx) = \chi(\pi(x) + \alpha) = \chi(\alpha) f(x)$, and *every* eigenfunction of T in $L^2(X, \mathcal{B}, \mu)$ is a multiple of one derived from a character. (Z, α) is unique up to isomorphism and is called the *Kronecker factor* of $(X, \mathcal{B}, \mu, T)$. The eigenvalues of T are $\{\chi(\alpha)\}_{\chi \in \hat{Z}}$, so that $Z \cong$ the dual group to the (discrete) group of eigenvalues of T. The system $(X, \mathcal{B}, \mu, T)$ is *weakly mixing* if and only if there are no eigenvalues other than 1 if and only if Z is the trivial one-element group. The discussion in this section will be vacuous in the case of weakly mixing systems.

We turn to processes. When we speak of an eigenfunction f we will assume $f \not\equiv 0$.

Definition 6.1. *A process* $(X, \mathcal{B}, \mu, T, \Phi)$ *is* Kronecker-complete *if it is ergodic and if every eigenfunction of* T *is proportional to some function in* Φ.

Note that for an ergodic system, if $Tf = \lambda f$ for a measurable f, it is easily seen that $|\lambda| = 1$ and that $|f(x)|$ is constant a.e., so that $f \in L^\infty(X, \mathcal{B}, \mu)$. Also note that $Tf_1 = \lambda f_1, Tf_2 = \lambda f_2$ implies that f_1/f_2 is invariant so that by ergodicity, f_1, f_2 are proportional. Thus a process is Kronecker-complete if Φ contains *some* eigenfunction for each eigenvalue. Under our standing hypothesis that $\mathcal{B}$ is a countably generated σ-algebra, the set of eigenvalues is at most countable. As a result we can always "complete" a non-Kronecker-complete process. The principal result in this section states that if a scheme is generic for a non-Kronecker-complete process, by augmenting the process and the scheme and passing to a subscheme, we will obtain a scheme generic for a Kronecker-complete process.

Theorem 6.2. *Let* $(\{J_l\},\{\xi_r^l\})$ *be generic for an ergodic process* $(X,\mathcal{B},\mu,T,\Phi)$. *Denote by* Λ *the subgroup of the unit circle* S^1 *consisting of eigenvalues of* T *on* $L^2(X,\mathcal{B},\mu)$. *We can find eigenfunctions* ψ_λ *for each* $\lambda \in \Lambda$ *and a subscheme* $(\{H_k\},\{\eta_r^k\})$ *so that setting* $\eta_\lambda^k(n) = \lambda^n$ *independent of* k *and letting* $\Psi = \{\psi_\lambda\}_{\lambda\in\Lambda}$, *the process* $(X,\mathcal{B},\mu,T,\Phi\cup\Psi)$ *will be Kronecker-complete, and the scheme* $(\{H_k\},\{\xi_r^k\}\cup\{\eta_\lambda^k\})$ *will be generic for* $(X,\mathcal{B},\mu,T,\Phi\cup\Psi)$.

In the weak mixing case we merely need to adjoin the function 1 to the process and to the scheme. In the general case we proceed by successively adjoining eigenfunctions, passing to a subarray at each stage. We will thus obtain a sequence of subarrays which is "decreasing" and a sequence $\Phi_n = \Phi \cup \{\Psi_{\lambda_1}, \Psi_{\lambda_2}, \dots, \Psi_{\lambda_n}\}$ of sets of functions with the corresponding $\{\eta_r^{(k)}\} \cup \{\eta_{\lambda_1}, \eta_{\lambda_2}, \dots, \eta_{\lambda_n}\}$ of representative time series. Our final scheme is obtained by choosing from successive schemes intervals that are "ε-h-generic" for the final process $(X,\mathcal{B},\mu,T,\Phi\cup\Psi)$ with $\varepsilon \searrow 0$, $h \nearrow \infty$. Such intervals will be found in the array for Φ_n with n sufficiently large.

Adjoining a single eigenfunction will also entail a procedure of successive approximation. We assume given a scheme $(\{J_l\},\{\xi_r^l\})$ generic for $(X,\mathcal{B},\mu,T,\Phi)$ and we wish to adjoin an eigenfunction for the eigenvalue λ. Fix an eigenfunction $f, Tf = \lambda f$, with $|f| = 1$. Since we have fixed the representative time series for the eigenfunction as η_λ where $\eta_\lambda(n) = \lambda^n$, the corresponding φ_λ to be adjoined will be some multiple cf, $|c| = 1$. Our task is to find subintervals of J_l that give better and better representation for the augmented $\Phi \cup \{cf\}$ in a sense analogous to ε-h-genericity (§2). In our procedure of successive approximation we can let c vary, since a subsequence will converge to a fixed value for which the intervals that have been found will still provide good representation. We form Φ^* from Φ as in §2, and express Φ^* as a union $\Phi^* = \bigcup \Phi_h^*$ of increasing finite subsets. Now $\{cf\}$ enters the picture and we say that the interval $J \subset J_l$ is "ε-h-m-generic" for $(X,\mathcal{B},\mu,T,\Phi\cup cf)$ if for every $\varphi \in \Phi_h^*$ and the corresponding time series ξ^l, and for a an integer with $0 \le a \le m$,

$$\left|\frac{1}{|J|}\sum_{n\in J}\xi^l(n)\lambda^{an} - \int_X \varphi\cdot c^a f^a d_\mu\right| < \varepsilon. \tag{5}$$

Note that for $a = 0$ this is ε-h-genericity. What will be shown for the proof of the theorem is the existence of ε-h-m-generic intervals inside J_l for large l for arbitrary ε, h, m, and putting these together we obtain the subscheme that is sought.

In establishing (5) we will use the following lemma.

Lemma 6.3. *Let* $\alpha_1, \alpha_2, \ldots, \alpha_N$ *be* N *complex numbers and form, for* $a, b = 0, 1, 2, \ldots$*, the averages*

$$u(a,b) = \frac{1}{N} \sum_{i=1}^{N} \alpha_i^a \overline{\alpha}_i^b.$$

There is a function $\delta(\varepsilon, p) > 0$ *for* $\varepsilon > 0$ *and* $p \in \mathbb{N}$ *so that if* $|u(a,b) - 1| < \delta(\varepsilon, p)$ *for* $0 \leq a, b \leq p$*, then* $\exists \beta$ *so that*

$$\frac{1}{N} \sum_{i=1}^{N} |\alpha_i - \beta|^{2p} < \varepsilon$$

Proof. We form the average

$$\begin{aligned} \frac{1}{N^2} \sum_{i,j=1}^{N} |\alpha_i - \alpha_j|^{2p} &= \frac{1}{N^2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_j)^p (\overline{\alpha}_i - \overline{\alpha}_j)^p \\ &= \frac{1}{N^2} \sum_{q=0}^{p} \sum_{q'=0}^{p} \sum_{i,j=1}^{N} (-1)^{2p-q-q'} \tbinom{p}{q} \tbinom{p}{q'} \alpha_i^q \, \overline{\alpha}_i^{q'} \, \alpha_j^{p-q} \, \overline{\alpha}_j^{p-q'} \\ &= \sum_{q,q'=0}^{p} (-1)^{2p-q-q'} \tbinom{p}{q} \tbinom{p}{q'} u(q,q') u(p-q,\, p-q') \end{aligned}$$

The latter expression is continuous in the $(p+1)^2$ expressions $\{u(q,q'),\ 0 \leq q,\ q' \leq p\}$ and we can evaluate it for $u(q,q') = 1$ by setting all $\alpha_i = 1$. Since the expression in question vanishes when $\alpha_i = 1$, it follows that we can find $\delta(\varepsilon, p) > 0$ so that the hypothesis of the lemma implies

$$\frac{1}{N} \sum_{j=1}^{N} \left(\frac{1}{N} \sum_{i=1}^{N} |\alpha_i - \alpha_j|^{2p} \right) < \varepsilon.$$

But this implies that for some index j the inside average is $< \varepsilon$, so with $\beta = \alpha_j$ we get the desired result. □

Proof of Theorem 6.2. We have seen that to prove the theorem we have to show the existence of long intervals J inside J_l for sufficiently large l, for which (5) is valid, where φ ranges over Φ_h^*, f is an eigenfunction $Tf = \lambda f$, and the $\xi^l(n)$ are the time series representing φ in the respective J_l, and the exponent "a" ranges from 1 to m.

Our assumption in Definition 2.5 that the functions of Φ generate the σ-algebra $\mathcal{B}$ for the process $(X, \mathcal{B}, \mu, T, \Phi)$ implies that linear combinations of functions in Φ^* will approximate any function in $L^p(X, \mathcal{B}, \mu)$ in the L^p-norm, for any p, $1 \leq p < \infty$. We wish to approximate f and for any $\varepsilon_1 > 0$ we can find σ in the linear space spanned by Φ^* with $||\sigma - f||_{L^q} < \varepsilon_1$ where $q = q(m) \geq 8$ will be made explicit further on. Taking appropriate combinations of the time series $\zeta^l(n)$ representing σ in the given scheme, we find that

$$\frac{1}{|J_l|}\sum_{n\in J_l}\left(\zeta^l(n)\right)^r\left(\overline{\zeta^l(n)}\right)^s\left(\zeta^l(n+k)\right)^t\left(\overline{\zeta^l(n+k)}\right)^u \longrightarrow \int\limits_X \sigma^r\overline{\sigma}^s T^k(\sigma^t\overline{\sigma}^u)d\mu. \tag{6}$$

We're going to apply Lemma 6.3 with $p = 2$ to the $N = K|J_l|$ numbers:

$$\alpha_{k,n} = \lambda^{-k}\zeta(n+k)\overline{\zeta(n)} \qquad 0 \le k \le K-1, \quad n \in J_l$$

where $\zeta = \zeta^l$. K will be arbitrary and l will be large. We have

$$u(a,b) = \frac{1}{|J_l|}\frac{1}{K}\sum_{k=0}^{K-1}\sum_{n\in J_l}\lambda^{(b-a)k}\zeta(n+k)^a\overline{\zeta(n+k)}^b\zeta(n)^b\overline{\zeta(n)}^a$$

When l is large this is close to $\int_X \frac{1}{K}\sum_{k=0}^{K-1}\lambda^{(b-a)k}\sigma^b\overline{\sigma}^a T^k(\sigma^a\overline{\sigma}^b)d\mu$. The latter expression will be within ε_2 of

$$\int\limits_X \frac{1}{K}\sum_{k=0}^{K-1}\lambda^{(b-a)k}f^b\overline{f}^a T^k(f^a\overline{f}^b)d\mu = \int\limits_X \frac{1}{K}\sum_{k=0}^{K-1}\lambda^{(b-a)k}f^{b-a}T^k(f^{a-b})d\mu = 1$$

where $\varepsilon_2 = \varepsilon_2(\varepsilon_1) \to 0$ as $\varepsilon_1 \to 0$, using the fact that σ is close to f in L^8 and the total exponent in the integrals above is $2a+2b \le 8$, and the fact that $T^k f = \lambda^k f$. Having chosen ε_1 sufficiently small, we find by Lemma 6.3 that for l large we can find β_l so that

$$\frac{1}{K}\frac{1}{J_l}\sum_{k,n}\left|\lambda^{-k}\zeta(n+k)\overline{\zeta(n)} - \beta_l\right|^4 < \varepsilon_0 \tag{7}$$

where ε_0 is given.

We wish to use (7) to estimate

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\left|\zeta(n+k)\overline{\zeta(n)}\zeta(n) - \lambda^k\beta_l\zeta(n)\right|^2 =$$
$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\left|(\lambda^{-k}\zeta(n+k)\overline{\zeta(n)} - \beta_l\right|^2|\zeta(n)|^2 \le \sqrt{\varepsilon_0}\theta_l$$

where $\theta_l^2 = \frac{1}{|J_l|}\sum_{n\in J_l}|\zeta(n)|^4 = \frac{1}{|J_l|}\sum_{n\in J_l}\zeta(n)^2\,\overline{\zeta(n)}^2$, and by (6), $\theta_l^2 \to \int|\sigma|^4 d\mu$ as $l \to \infty$. Since $\|\sigma - f\|_4 < \varepsilon_1$ the latter expression is $< (1+\varepsilon_1)^4$ and we can assume this ≤ 4. We get for large l

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\left|\zeta(n+k)|\zeta(n)|^2 - \lambda^k\beta_l\zeta(n)\right|^2 < 2\sqrt{\varepsilon_0}. \tag{8}$$

Finally we wish to use this to estimate

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\left|\zeta(n+k) - \lambda^k\beta_l\zeta(n)\right|^2$$

and for this we need an estimate of

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\Big|\zeta(n+k)|\zeta(n)|^2-\zeta(n+k)\Big|^2. \tag{9}$$

As $l\to\infty$, (9) approaches

$$\int\Big(|\sigma|^4\,T^k|\sigma|^2-2|\sigma|^2\,T^k\,|\sigma|^2\;+T^k|\sigma|^2\Big)\,d\mu. \tag{10}$$

The corresponding expression for f instead of σ vanishes so that for some C, the expression in (10) is bounded by $C\varepsilon_1$, and the same will be true for (9) when l is large. Combining this with (8) gives

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\Big|\zeta(n+k)-\lambda^k\beta_l\zeta(n)\Big|^2<\varepsilon_3=\varepsilon_3(\varepsilon_1)$$

for large l, where $\varepsilon_3(\varepsilon_1)\to 0$ for $\varepsilon_1\to 0$.

Using the Hilbert space inequality

$$||u||^2-||v||^2\le\Big(||u||+||v||\Big)\,||u-v||$$

we find for large l

$$\left|\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}|\zeta(n+k)|^2-\frac{1}{K}\frac{1}{|J_l|}\,|\beta_l|^2\,\sum_{k,n}|\zeta(n)|^2\right|\le C'\sqrt{\varepsilon_3}$$

from which it follows that $|\beta_l|\to 1$. To summarize the foregoing, we have shown that for any $\varepsilon>0$ we can find a function σ with time series $\zeta^l(n)$ and γ_l with $|\gamma_l|=1$ so that for l sufficiently large, and any K,

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\Big|\zeta^l(n+k)-\lambda^k\gamma_l\zeta^l(n)\Big|^2<\varepsilon.$$

To apply this to (5) we let $1\le a\le m$ and we estimate for a time series $\xi^l(n)$

$$\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}\Big|\Big(\zeta^l(n+k)\Big)^a-\lambda^{ak}\gamma_l\Big(\zeta^l(n)\Big)^a\Big|\;\Big|\xi^l(n+k)\Big| \tag{11}$$

Writing $x^a-y^a=(x-y)(x^{a-1}+x^{a-2}y+\cdots+y^{a-1})$ we obtain for large l that the expression in (11) is bounded by $M\sqrt{\varepsilon}$ where

$$M^2=a\Bigg(\sum_{j=0}^{a-1}\frac{1}{K}\frac{1}{|J_l|}\sum_{k,n}|\zeta^l(n+k)|^{2(a-j-1)}|\zeta^l(n)|^{2j}|\xi^l(n+k)|^2\Bigg)$$

If ξ^l represents the function φ, the limit of the foregoing expression, as $l \to \infty$, is

$$\frac{a}{K} \sum_{k=0}^{K-1} \int T^k |\sigma|^{2(a-j-1)} |\sigma|^{2j} T^k |\varphi|^2 d\mu,$$

and provided $q(m) \geq 2m+2$ with $\|\sigma - f\|_{L^q} < 1$, the expression in (11) will be bounded by $C' \|\varphi\|_{L^q} \sqrt{\varepsilon},\ C' = C'(m)$.

In all the estimates for averages over $0 \leq k < K,\ n \in J_l$, if the overall average is $< \theta$, then for at least half of the $n \in J_l$, the average over k cannot exceed 2θ. For large l, we let $N_l \subset J_l$ consist of the n with $\{n, n+1, \ldots, n+K-1\} \subset J_l$ and

$$\left| \frac{1}{K} \sum_{k=0}^{K-1} \left(\left(\zeta^l(n+k)\right)^a - \lambda^{ak} \left(\gamma_l \zeta^l(n)\right)^a \right) \xi^l(n+k) \right| < 2C' \|\varphi\|_{L^q} \sqrt{\varepsilon}. \quad (12)$$

We now refer to Theorem 2.9 applied to the functions $\sigma_\varphi^a,\ 1 \leq a \leq m$, $\varphi \in \Phi_h^*$ which are in the linear span of Φ^*. These functions are represented in the given scheme by $\left(\zeta^l(n)\right)^a \xi^l(n)$, and with $\delta > 0$ given, there will be a K so that for sufficiently large l, the inequalities

$$\left| \frac{1}{K} \sum_{k=0}^{K-1} \left(\zeta^l(n+k)\right)^a \xi^l(n+k) - \int \sigma^a \varphi d\mu \right| < \delta \quad (13)$$

hold for most $n \in J_l$ provided $|J_l| \gg K$. This implies that (12) and (13) will hold simultaneously for most $n \in N_l$ for which we will then have

$$\left| \frac{1}{K} \sum_{k=0}^{K-1} \lambda^{ak} \left(\gamma_l \zeta^l(n)\right)^a \xi^l(n+k) - \int \sigma^a \varphi d\mu \right| < \delta + 2C' \|\varphi\|_{L^q} \sqrt{\varepsilon}.$$

Set $c_{l,n} = \lambda^n \gamma_l^{-1} \zeta^l(n)^{-1}$ and we can write

$$\left| \frac{1}{K} \sum_{k=0}^{K-1} \lambda^{a(n+k)} \xi^l(n+k) - c_{l,n}^a \int \sigma^a \varphi d\mu \right| < |c_{l,n}|^a \left(\delta + 2C' \|\varphi\|_{L^q} \sqrt{\varepsilon}\right) \quad (14)$$

We write $n \in N_l'$ if (14) is valid.

$$\left| \frac{1}{K} \sum_{k=0}^{K-1} \lambda^{a(n+k)}\ \xi^{(l)}(n+k) - c_{l,n}^a \int_X f^a \varphi d\mu \right| < \quad (15)$$
$$|c_{l,n}|^a \left(\delta + 2C' \|\varphi\|_{L^q} \sqrt{\varepsilon}\right) + |c_{l,n}|^a c'' \|\varphi\|_{L^{m+1}} \|\sigma - f\|_{L^{m+1}}$$

If J is the interval $\{n, n+1, \ldots, n+k-1\}$ then (15) has the form (5) if the right hand side can be made small and if $|c_{l,n}|$ is close to 1. All this can be achieved by choosing σ with $\|\sigma - f\|_{L^q}$ small, and finding $n_l \in J_l$ for which (15) holds with $|c_{l,n}| = |\gamma_l|^{-1}|\zeta^l(n)|^{-a}$ close to 1. The domain of n is N_l' which depends on ζ^l, but $|N_l'|/|J_l|$ is bounded from below. It suffices to show that by choosing $\|\sigma - \delta\|_{L^q}$ small we will have (for ζ^l representing σ) $\Big||\zeta^{(l)}(n)| - 1\Big| < \theta$ for a preassigned $\theta > 0$ for most $n \in J_l$. But this follows from the fact that

$$\frac{1}{K} \sum_{n \in J_l} \left(|\zeta^l(n)|^2 - 1\right)^2 \longrightarrow \int \left(|\sigma|^2 - 1\right)^2 d\mu$$

as $l \to \infty$ and the latter expression is small if $\|\sigma - f\|$ is small. With this we have completed the proof of Theorem 6.2. □

Corollary 6.4. *If an ergodic process is Kronecker-complete, it has a generic scheme whereby eigenfunctions are represented by the time series $c_\lambda \lambda^n$ for all intervals of the array $\{J_l\}$.*

Suppose now that we have a generic scheme for a Kronecker-complete process, $(X, \mathcal{B}, \mu, T, \Phi)$ and let $\Lambda \subset S^1$ be the group of eigenvalues of the process. If we identify the Kronecker factor of $(X, \mathcal{B}, \mu, T)$ with $Z = \hat{\Lambda}$ we can define a *canonical map* $\pi: X \to Z$. Namely for $\lambda \in \Lambda$ there is a unique eigenfunction φ_λ on X with $T\varphi_\lambda = \lambda\varphi_\lambda$, and which is represented in the scheme by $\eta_\lambda(n) = \lambda^n$. We set $\alpha \in Z = \hat{\Lambda}$ to correspond to the inclusion map of $\Lambda \to S^1 : \alpha(\lambda) = \lambda$. Notice that since $\eta_{\lambda_1\lambda_2} = \eta_{\lambda_1}\eta_{\lambda_2}$ we will have $\varphi_{\lambda_1\lambda_2} = \varphi_{\lambda_1}\varphi_{\lambda_2}$. This means that for a.e. $x \in X$, $\varphi_{\lambda_1\lambda_2}(x) = \varphi_{\lambda_1}(x)\varphi_{\lambda_2}(x)$ so that if we define $\pi(x)(\lambda) = \varphi_\lambda(x)$, then for a.e. $x, \pi(x) \in \hat{\Lambda} = Z$. Moreover $\pi(Tx)(\lambda) = \varphi_\lambda(Tx) = \lambda\varphi_\lambda(x) = \alpha(\lambda)\pi(x)(\lambda) = (\alpha + \pi(x))(\lambda)$; so $\pi(Tx) = \pi(x) + \alpha$. The mapping π is measurable since all φ_λ are measurable, and so the foregoing gives an explicit map of X to its Kronecker factor. This map will play a role in §7.

Note that for $\lambda \in \Lambda$, the eigenfunction φ_λ on X can be identified with $\chi \circ \pi$, where χ is the character on Z given by $\chi(z) = z(\lambda)$ where Z is identified with $\hat{\Lambda}$, since $\chi(\pi(x)) = \pi(x)(\lambda) = \varphi_\lambda(x)$ by definition of π. Since the time series representing φ_λ is $\lambda^n = \chi(n\alpha)$, we conclude:

Proposition 6.5. *Given a scheme generic for a Kronecker-complete process $(X, \mathcal{B}, \mu, T, \Phi)$, if π is the canonical map of X to its Kronecker factor (Z, α) then for any continuous function ψ on Z, $\psi \circ \pi$ can be adjoined to Φ, and it will be represented by the time series $\{\psi(n\alpha)\}$.*

Proof. ψ can be approximated uniformly by linear combinations of $\{\varphi_\lambda\}$. □

7 Weighted Ergodic Averages for Kronecker-complete Processes

Let $(X, \mathcal{B}, \mu, T, \Phi)$ be a Kronecker-complete process and $(\{J_l\}, \{\xi_r^l\})$ a generic scheme. We shall show how to evaluate L^2-limits of weighted ergodic averages

$$\frac{1}{|J_l|}\sum_{n\in J_l}\xi^l(n)T^n f$$

for $f\varepsilon L^2(X,\mathcal{B},\mu)$ and ξ^l representing a function $\varphi\in\Phi$. By our assumption $(X,\mathcal{B},\mu,T)$ is ergodic so that $\frac{1}{N}\sum_{n=0}^{N-1}T^n f\to\int f d\mu$ in L^2. Since T is a contraction we can write

$$\frac{1}{|J_l|}\sum_{n\in J_l}T^n f\longrightarrow\int f d\mu$$

for any array $\{J_l\}$. This will be generalized for processes that are Kronecker-complete, except that the limits taken are weak L^2-limits.

Recall from §6 the notion of Kronecker factor and the canonical map $\pi: X\to Z$ where Z is a compact abelian group and $\pi(Tx)=\pi(x)+\alpha$. All eigenfunctions on X are, up to constant multiples, of the form $\chi\circ\pi$ where χ is a character on Z. The set of all functions in $L^2(X,\mathcal{B},\mu)$ of the form $\psi\circ\pi,\psi\in L^2(Z)$ form a subspace that is spanned by eigenfunctions. If $f\in L^2(X,\mathcal{B},\mu)$ we denote by $E(f|Z)$ the unique function in $L^2(Z)$ so that $E(f|Z)\circ\pi$ denotes the orthogonal projection of f to the subspace $L^2(Z)\circ\pi$. $E(f|Z)=0\Leftrightarrow f$ is orthogonal to all eigenfunctions in $L^2(X,\mathcal{B},\mu)$. We will make use of an operation on $L^1(Z)$ related to (but not the same as) convolution:

$$f_1\Box f_2(z)=\int_Z f_1(z+u)f_2(u)du$$

Proposition 7.1. *Let* $(\{J_l\},\{\xi_r^l\})$ *be generic for the Kronecker-complete process* $(X,\mathcal{B},\mu,T,\Phi)$, *let* $f\in L^2(X,\mathcal{B},\mu)$, *and let* $\varphi\in\Phi$ *be represented by the time series* ξ^l. *Then*

$$\frac{1}{J_l}\sum_{n\in J_l}\xi^l(n)T^n f\stackrel{w}{\longrightarrow}\Big[E(f|Z)\Box E(\varphi|Z)\Big]\circ\pi \tag{16}$$

where $\stackrel{w}{\longrightarrow}$ *signifies weak convergence in* $L^2(X,\mathcal{B},\mu)$.

Proof. It suffices to consider two cases: (a) $E(f|Z)=0$, (b) f is an eigenfunction.

In the first case, for any g in $L^2(X,\mathcal{B},\mu)$, the sequence $\{\int T^n f\cdot g d\mu\}$ satisfies

$$\frac{1}{N}\sum_{k=n+1}^{n+N}\left|\int T^k f\cdot g d\mu\right|^2\underset{N\to\infty}{\longrightarrow}0$$

uniformly in n, so that the left hand side of (16) goes to 0 weakly, and the proposition is verified. We turn to case (b) with $f=\varphi_\lambda$. To $\lambda\in\Lambda$ we associate the character χ on $\hat{\Lambda}$ with $\chi(z)=z(\lambda)$. Then $\chi\circ\pi(x)=\pi(x)(\lambda)=\varphi_\lambda(x)=f(x)$, and $E(f|Z)=\chi$. In this case the right hand side of (16) is $[\chi \Box E(\varphi|Z)] \circ \pi = \left(\int\limits_Z E(\varphi|Z)\chi dz\right)\chi \circ \pi$. We evaluate the left hand side of (16):

$$\frac{1}{|J_l|}\sum_{n\in J_l} \xi^l(n)T^n f = \frac{1}{|J_l|}\sum_{n\in J_l} \lambda^n \xi^{(l)}(n) f$$

which by genericity converges to $\left(\int \varphi_\lambda \varphi d\mu\right) f$. Since $\varphi_\lambda \in L^2(Z) \circ \pi$, we can replace φ by its projection to this subspace which is $E(\varphi|Z) \circ \pi$. Since $\varphi_\lambda = \chi \circ \pi$ we now have

$$\int\limits_X \varphi_\lambda \cdot \varphi d\mu = \int\limits_Z \chi E(\varphi|Z) dz$$

and since $f = \chi \circ \pi$, this proves the proposition. ☐

8 A Condition for PW-Bohr$_0$

We know from Theorem 3.5 that if $d^*(S) > 0$ for a subset $S \subset \mathbb{Z}$, then $S - S$ is PW-Bohr$_0$. We can rephrase this as saying that if for each $s \in S$, $S - s \cap B = \emptyset$ for a subset $B \subset \mathbb{Z}$, then the complement of B is PW-Bohr$_0$. In this section we show that it will suffice for this conclusion that $d^*\Big((S - s) \cap B\Big) = 0$ for each $s \in S$. In §9 we'll see how this leads to a proof of Theorem II.

Proposition 8.1. *Let $A \subset \mathbb{Z}$ and $B = \mathbb{Z}\backslash A$ and let $S \subset \mathbb{Z}$ with $d^*(S) > 0$. If for every $s \in S$, $d^*\Big((S - s) \cap B\Big) = 0$, then A is a PW-Bohr$_0$ set.*

Proof. Let $\{J_l\}$ be an array with $\frac{|J_l \cap S|}{|J_l|} \to \beta > 0$. Set $\xi_1^l(n) = 1_A(n)$, $\xi_2^l = 1_B(n), \xi_3^l(n) = 1_S(n)$ and consider the scheme $(\{J_l\}\ \{\xi_1^l, \xi_2^l, \xi_3^l\})$. By Theorem 2.9 we can find a subscheme generic for an ergodic process $(X, \mathcal{B}, \mu, T, \Phi)$ where Φ includes $\varphi_1, \varphi_2, \varphi_3$ which are respectively represented by $\xi_1^l, \xi_2^l, \xi_3^l$. By the scholium to Theorem 2.9 we can assume φ_3 is not a.e. 0. Since $(\xi_i^{(l)})^2 = \xi_i^{(l)}$ we find $\varphi_i^2 = \varphi_i$ a.e. and so φ_i take values 0, 1. We write $\varphi_1 = 1_{\tilde{A}}$, $\varphi_2 = 1_{\tilde{B}}$, $\varphi_3 = 1_{\tilde{S}}$ with $\tilde{A}, \tilde{B}, \tilde{S} \subset X$, $\mu(\tilde{S}) > 0$, and $\tilde{A} \cup \tilde{B} = X$. Using Theorem 6.2 we can also assume that the process $(X, \mathcal{B}, \mu, T, \Phi)$ is Kronecker-complete and that the eigenfunctions $\{\varphi_\lambda\}$ of the process are represented by time series $\eta_\lambda(n) = \lambda^n$. We will also make use of the canonical map $\pi: X \to Z$, where (Z, α) is the Kronecker factor of $(X, \mathcal{B}, \mu, T)$.

We now apply Proposition 7.1 to this subscheme generic for the Kronecker-complete process with $\varphi_1, \varphi_2, \varphi_3 \in \Phi$, and where we again denote the array of intervals by $\{J_l\}$. We will take $f = \varphi = 1_{\tilde{S}} = \varphi_3$ which is represented by $\xi_3^l(n) = 1_S(n)$. We conclude that in the weak L^2-topology,

$$\frac{1}{|J_l|}\sum_{n\in J_l} 1_S(n)T^n 1_{\tilde{S}} \longrightarrow (f\Box f)\circ\pi \tag{17}$$

where $f = E(1_{\tilde{S}}|Z)$. The function f is bounded and non-negative with $\int f dz = \mu(\tilde{S}) > 0$ so it is non-trivial. We note that since $f \in L^\infty(Z)$, the function $F = f\Box f$ is continuous on Z.

We turn now to the hypothesis that $d^*\Big((S-s)\cap B\Big) = 0$ for $s \in S$. This implies that

$$\frac{1}{|J_l|}\sum_{n\in J_l} 1_B(n)1_S(n+s) \longrightarrow 0$$

or

$$\int 1_{\tilde{B}} T^s 1_{\tilde{S}} d\mu = 0.$$

In particular, averaging over $s \in S$:

$$\frac{1}{|J_l|}\sum_{n\in J_l} 1_S(n)\int T^n 1_{\tilde{S}} 1_{\tilde{B}}\, d\mu \longrightarrow 0. \tag{18}$$

But by (17), the limit in (18) is

$$\int F\circ\pi\cdot 1_{\tilde{B}}\, d\mu \tag{19}$$

and so the latter integral vanishes. We again apply the generic scheme where according to Corollary 6.4, $F\circ\pi$ is represented by $\{F(n\alpha)\}$, a non-negative almost periodic sequence with

$$\frac{1}{|J_l|}\sum_{n\in J_l} F(n\alpha) \longrightarrow \int F dz > 0$$

Since the integral in (19) vanishes we can write

$$\frac{1}{|J_l|}\sum_{n\in J_l} F(n\alpha)1_B(n) \longrightarrow 0$$

Let H be the Bohr_0 set for which $F(n\alpha) > \delta$ where $\delta > 0$ is chosen so that H is non-empty. Then

$$\frac{\sum_{n\in J_l} 1_H(n)1_B(n)}{\sum_{n\in J_l} 1_H(n)} \longrightarrow 0$$

whence

$$\frac{\sum_{n\in J_l} 1_{H\cap A}(n)}{\sum_{n\in J_l} 1_H(n)} \longrightarrow 1.$$

This implies that there are arbitrarily long intervals $L_l \subset J_l$ for which $H\cap L_l = H\cap A\cap L_l \subset A$. Hence $H\cap\bigcup L_l \subset A$ from which it follows that A is PW-Bohr_0. This proves Proposition 8.1. $\Box$

9 Application to Δ_r^*-sets

We shall apply the foregoing results to prove Theorem II of §1. We recall that a subset $A \subset \mathbb{Z}$ is a Δ_r^*-set, $r = 2, 3, \ldots$ if for distinct numbers $x_1, x_2, \ldots, x_r$, some difference $x_j - x_i$, $i < j$ belongs to A. More generally we will need

Definition 9.1. *If $S \subset \mathbb{Z}$ we shall write $A \in \Delta_r^*(S)$ if for $x_1, x_2, \ldots, x_r \in S$, $x_i \neq x_j$ for $i \neq j$, there exists $i < j$ with $x_j - x_i \in A$.*

In the sequel, A and B denote complementary sets in $\mathbb{Z}$, $B = \mathbb{Z} \backslash A$. If $0 \in B$ we denote by B' the set $B \backslash \{0\}$.

Lemma 9.2. *The following are equivalent for a set $S \subset \mathbb{Z}$:*

(a) $A \in \Delta_{r+1}^*(S)$

(b) $A \in \Delta_r^*\Big(B' \cap (S - s)\Big)$ *for every $s \in S$.*

Proof. (a) ⇒ (b): Suppose $x_1, x_2, \ldots, x_r \in B' \cap (S - s)$. Form the $(r+1)$-tuple $s, s + x_1, s + x_2, \ldots, s + x_r$ and apply (a). (b) ⇒ (a): Let $x_0, x_1, x_2, \ldots, x_r$ be distinct elements in S. If $\{x_1 - x_0,\ x_2 - x_0, \ldots, x_r - x_0\}$ doesn't meet A, then this is an r-tuple in $B' \cap (S - x_0)$ and we can apply (b). □

We recall Theorem II:

Theorem II. *For any $r = 2, 3, \ldots$, if A is a Δ_r^*-set then A is a PW-Bohr$_0$.*

Proof. We assume A is not PW-Bohr$_0$. By Proposition 8.1 this will imply that whenever $d^*(S) > 0$ there must be some $s \in S$ with $d^*\Big(B \cap (S - s)\Big) > 0$. This will give us an inductive procedure to obtain sets S_i with $d^*(S_i) > 0$. Start with $S_0 = \mathbb{Z}$ and we find $d^*(B) > 0$. Set $S_1 = B'$, there exists $s_1 \in S_1$ with $d^*(B \cap (S_1 - s_1)) > 0$. Set $S_2 = B' \cap (S_1 - s_1)$ and continue with $S_{k+1} = B' \cap (S_k - s_k)$, $s_k \in S_k$. Now apply the foregoing lemma. $A \in \Delta_r^* \Leftrightarrow A \in \Delta_r^*(\mathbb{Z}) \Rightarrow A \in \Delta_{r-1}^*\Big(B' \cap (\mathbb{Z} - s_0)\Big) = \Delta_{r-1}^*(S_1) \Rightarrow A \in \Delta_{r-2}^*\Big(B' \cap (S_1 - s_1)\Big) = \Delta_{r-2}^*(S_2) \Rightarrow \cdots$ We continue with $A \in \Delta_{r-k}^*(S_k)$ for $k = 0, 1, \ldots, r-2$. Finally $A \in \Delta_2^*(S_{r-2})$. At each stage we have $d^*(S_k) > 0$. But $d^*(S_{r-2}) > 0 \Rightarrow S_{r-2} - S_{r-2}$ is PW-Bohr$_0$; and $A \in \Delta_2^*(S_{r-2}) \Rightarrow A \supset S_{r-2} - S_{r-2}$. This proves the theorem. □

References

[Ber96] V. Bergelson, Ergodic Ramsey theory – an update in *Ergodic Theory of $\mathbb{Z}^d$-actions*, London Math. Soc. Lecture Notes Series 228 (1996), 1-61.

[BH96] V. Bergelson, I. J. Håland, Sets of recurrence and generalized polynomials, Convergence in Ergodic Theory and Probability (Columbus, OH, 1993) 91–110, *Ohio State University Math. Res. Inst. Publ.* 5, de Gruyter, Berlin, 1996.

[BFM96] V. Bergelson, H. Furstenberg, R. McCutcheon, IP-sets and polynomial recurrence, *Ergodic Theory and Dynamical Systems* **16** (1996) no. 5, 963–974.

[Bilu97] Y. Bilu, Addition of sets of integers of positive density, *Journal of Number Theory* **64** (1997) 233-275.

[B-M86] A. Bertrand-Mathis, Ensembles intersectifs et recurrence de Poincaré, *Israel J. Math.* **55** (1986), 184–198.

[Bog39] N. Bogoliouboff, Sur quelques propriétés arithmetiques des presque periodes, *Ann. Chaire Phys. Math. Kiev* **4** (1939), 195–205.

[Bro71] T. C. Brown, An interesting combinatorial method in the theory of locally finite semigroups, *Pacific Journal of Math.* **36** (1971), 285–289.

[Føl54] E. Følner, Generalization of a theorem of Bogoliouboff to topological abelian groups, *Mathematica Scandinavica* **2** (1954), 5–19.

[Fur67] H. Furstenberg, Disjointness in ergodic theory, minimal sets, and a problem in diophantine approximation, *Math. Systems Theory* **1** (1967), 1–49.

[Fur81] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton, Princeton Univ. Press, 1981.

[GRS80] R. L. Graham, B. L. Rothschild, J. H. Spencer, *Ramsey Theory*, John Wiley, New York (1980).

[Hel83] H. Helson, *Harmonic Analysis*, Addison-Wesley Reading, MA, (1983).

[Jin02] R. Jin, The sumset phenomenon, *Proc. Amer. Math. Soc.* **130** (2002), no. 3, 855–861.

[Kre85] U. Krengel, *Ergodic Theorems*, De Gruyter Studies in Mathematics, 6, Walter de Gruyter & Co., Berlin, (1985) viii+357.

[Kříž87] I. Kříž, Large independent sets in shift-invariant graphs: solution of Bergelson's problem, *Graphs Combin.* **3** (1987) no. 2, 145–158.

[Ruz82] I. Ruzsa, Uniform distribution, positive trigonometric polynomials and difference sets, *Seminar on Number Theory*, 1981/1982, Exp. No. 18, 18 pp. Univ. Bordeaux I, Talence, 1982.

[Sár78] A. Sárközy, On difference sets of sequences of integers, III, *Acta Math. Acad. Sci. Hungar.* **31** (1978) no. 3-4, 355–386.

A Generalization of Conway Number Games to Multiple Players

Christopher Cunningham* and Igor Kriz*

Department of Mathematics, University of Michigan, 2074 East Hall, 530 Church Street, Ann Arbor, MI 48109-1043, USA
ccccc@umich.edu, ikriz@math.lsa.umich.edu

Summary. We define an analogue of the the concept of J. H. Conway's number games for games of multiple players. We define the value of such number game as an element of a vector space over the Conway field. We interpret the value in terms of the strategy of the game, and prove that all possible values in the vector space can occur.

AMS Subject Classification. 91A06, 91A46, 91A70.
Keywords. Multiple player games, deterministic games, number games, strategy, Conway field.

1 Introduction

There are many different mathematical meanings of the word 'game'. Regardless of the kind of games we consider, people agree that games of n players are much more difficult to understand for $n > 2$ than for $n \leq 2$. In this paper, we consider deterministic 'combinatorial' games, i.e. games where each player in each position has a well defined set of moves, which, in a fixed way, change the position to another position (in fact, it is clear that there is no point in distinguishing between positions and games, so we can substitute the word 'game' for the word 'position' everywhere). For some recent work on combinatorial games, see [Now02]. The main result of this paper is to analyze a certain, very special, class of combinatorial games for multiple players.

The definition given above, of course, describes only the 'static' aspect of the rules of a game. The 'dynamic' aspects refer to how the game is actually played. A play by play sequence of moves in a game will be called a 'match'. The dynamical rules of matches which we will consider specify a certain order of the set of players; the players shall move repeatedly in the same order of play until a certain player cannot move, at which point the match shall end.

* The authors were supported by NSF grant DMS 0305853.

The player who cannot move shall then be declared the loser of the match. We shall consider only games where there is no possibility of infinite matches.

Even for such deterministic games, however, it is difficult to make any conclusions about the course of matches for $n > 2$. The reason is that unlike the case of $n = 2$, there is no natural order of preference of the outcomes of the game from the point of view of the i'th player. While the i'th player obviously prefers not to lose, there is no natural reason why he should a priori prefer one particular other player to lose. Yet, such preferences will determine strategies, and ultimately the outcome of a match. Preferences can even change throughout the course of a match. Thus, it is usually said that few strategic conclusions about deterministic games of $n > 2$ players can be made without introducing non-deterministic concepts, perhaps even non-mathematical concepts (e.g. psychology).

The purpose of this paper is to look at a certain very special class of deterministic games of n players, for which certain strategical conclusions can be made in a rigorous mathematical setting, without introducing outside concepts. The motivation for introducing our particular class of games is that they generalize the 'number games' for 2 players from J. H. Conway's famous book [Con01]. For this reason, we shall call this class of games *number games for n players*.

Conway number games for 2 players are games to which one can assign a value which is a 'number'. Here the word number means element of a certain ordered field, known as the Conway field $\mathcal{C}$ (also known as the surreal numbers [Con01, Con72, Knu74]). The Conway field contains, among other things, all ordinal numbers, as well as any other ordered field: it is a foundational-level object of set theory (in fact, Conway introduces his own approach to formal set theory based on number games in [Con01]).

However, this is not the aspect of number games which we will be most interested in. Rather, the main point of number games is that 'no player can possibly improve his own position by making a move'. This, of course, needs precise definition, (in particular in reference to the word 'improve'). Definitions will be provided later. We shall, however, remark here that one could adopt the point of view that for this reason, number games are generally strategically uninteresting, since the very meaning of strategy is being able to take advantage of one's own move in the best possible way. In number games, one always makes one's own position worse by moving. One can also take another point of view; in some sense, number games are *pure consumer models*: we can think of moving in a number game as using up one's assets (resources). The player who uses up his resources first dies.

With this in mind, it becomes interesting to try to define analogues of Conway number games for n players, and analyze what, if any, strategic conclusions one can make for such games. In this paper, we do give one possible definition of such number games for n players. The set of (equivalence classes of) such games is an $n-1$-dimensional vector space over the Conway field. It is somewhat surprising the set of number games of n players has many of the formal properties of number games of 2 players. Also, we shall be able to make certain strategic conclusions for number games: in particular, each match will have a well defined 'loser' who can always be defeated if all the other players act 'in concert'.

This paper is organized as follows: In the next section, we shall present basic definitions and facts about games and matches which do not involve numbers. In Section 3, we shall introduce number games, and prove what we can say about their strategic analysis. Section 4 contains, in some sense, our hardest result, namely constructing number games of n players with any given value. The paper has two appendices. In Appendix 1 (section 4), we draw some diagrams visualizing our concepts for games of three players. This may be helpful to the reader in understanding what we mean. In Appendix 2 (section 4), we show why our definition of number game cannot be simplified in one obvious way.

2 Games and Matches

In this paper, a *game with set T of players* is defined recursively as follows:

1. The empty set $\emptyset$ is a game (also called 0).
2. If G_i are sets of games for all $i \in T$, then the tuple $G = (G_i)_{i \in T}$ is a game. (G_i is the set of possible moves of the player i in the game G; a game is identified with its initial position.)
3. Every game can be obtained by 1, 2 in a possibly transfinite number of steps. (This means that for every ordinal α, we have an "α'th generation of games", the 0'th generation being the empty game. For a game G of the α'th generation, every element of every G_i must be a game of generation $< \alpha$.)

Obviously, only the cardinality of the set T matters. We shall mostly consider the case $T = \{1, \ldots, n\}$ (in which case we shall simply speak of *games of n players*), but it is useful to allow other T's, notably $T \subset \{1, \ldots, n\}$.

We shall now introduce our main strategic concept for games of n players. It is important to notice that this concept does not involve dynamic aspects of games, i.e. matches.

Specifically, we shall inductively define

$$G <_S 0$$

for a non-empty set $S \subseteq T$ if the following conditions hold:

1. If $i \in S$ then for all $H \in G_i$, $H <_{\{i\}} 0$.
2. If $i \in S$ and $j \notin S$ then there exists an $H \in G_j$ and a set U with $i \in U \subseteq S \cup \{j\}$, such that $H <_U 0$.

We shall write

$$G \sim 0$$

if $G <_T 0$. The set S will be called the *set of possible losers of the game G*. We shall justify this terminology at the end of this section.

When working with games of n players, we shall use a notational convention analogous to that established in Conway's book [Con01], and denote a "general member" of the set G_i by G^i. Thus, for example, instead of referring to something that is true for *all* $H \in G_i$, we instead refer to something that is true for *all* G^i.

The members of G_i, or using the new convention, *the* G^i, are referred to as *i's options* in the game G.

In this notation, the above definition reads as follows:

Inductively define $G <_S 0$ for a non-empty set $S \subseteq T$ if the following conditions hold:

1. If $i \in S$ then all $G^i <_{\{i\}} 0$.
2. If $i \in S$ and $j \notin S$ then there exists a G^j and a set U with $i \in U \subseteq S \cup \{j\}$, such that $G^j <_U 0$.

Remark. When we are considering games of two players, we are in the context of Conway [Con01]. There, players are denoted by L and R, so $T = \{L, R\}$. Conway notes that it is easy to prove that all games of 2 players are of one of the following types: $0, L, R, F$. For 0, the first player loses, in F, the first player wins, in L (resp. R) the player L (resp. R) wins no matter whose move it is. In the above notation, $S = \{L, R\}$ for type 0, $S = R$ for type L, $S = L$ for type R and S does not exist for type F. The proof is left to the reader as an exercise.

Lemma 2.1. *For each game G, there exists at most one S such that $G <_S 0$.*

Proof. Induction: Note that

$$0 <_S 0 \text{ if and only if } S = T.$$

Assume the statement true for all G^i for all $i \in T$. Then if

$$G <_S 0,$$

note that $i \notin S$ if and only if there exists a G^i such that it is not true that $G^i <_{\{i\}} 0$, which uniquely characterizes S. ◻

We now explain the dynamical significance of $G <_S 0$. To this end, we must define matches. Assume now that T is finite, and that we have a bijection

$$\sigma : \{1, \dots, n\} \to T.$$

Such bijection will be called an *order of play.* A match according to the order of play σ is a sequence of games

$$(G(j))_{j=1,\dots,N}$$

where $G(1) = G$,

$$G(j+1) \in G(j)_{k(j)} \text{ where } k(j) = \sigma(j'),\ j' \equiv j \mod n,$$

$$G(N)_{k(N)} = \emptyset.$$

Then $k(N)$ is called *the loser of the match.* Note that

$$(G(j))_{j=2,\dots,N}$$

is a match according to the order of play σ' where

$$\sigma'(j) = \sigma(j+1) \text{ for } j < n,$$

$$\sigma'(n) = \sigma(1).$$

Note also that by part 3 of our definition of game, it is impossible to have an infinite match, i.e. an infinite sequence satisfying the properties of a match without the N. (Proof: induction.)

We now define inductively our main dynamic strategic concept. A player i is called the *loser of a game G according to the order of play σ* if

1. If $\sigma(1) = i$ then i is the loser of all its options G^i according the order of play σ'.
2. If $\sigma(1) \neq i$, then there exists a $G^{\sigma(1)}$ such that i is the loser of $G^{\sigma(1)}$ according to the order of play σ'.

Intuitively speaking, this means that i will lose any match according to the order of play σ, provided that all the other players act "in concert".

Proposition 2.2. *Suppose $G <_S 0$ and suppose that σ is any order of play. Let j be minimal such that $\sigma(j) \in S$. Then $\sigma(j)$ is the loser of the game G according to the order of play σ.*

Proof. Induction. If $\sigma(1) = i$, then always $G^i <_{\{i\}} 0$, so the induction hypothesis applies. If $\sigma(1) \neq i$, then there exists a $G^{\sigma(1)}$ such that $G^{\sigma(1)} <_U 0$ for some $i \in U \subseteq S \cup \{\sigma(1)\}$. Note that i satisfies the induction hypothesis with G replaced by $G^{\sigma(1)}$, and σ replaced by σ'. □

With this new dynamic significance applied to our previous definitions, the definitions can be formulated in a more intuitive way. If we find a set $S \subseteq T$ with $G <_S 0$, then the set S is the set of players who, for some order of play σ, would definitely lose the game if the others acted in concert. This is the reason the set S can be thought of as the set of possible losers of the game G. This can yield intuitive versions of 1 and 2 of the previous definition. $G <_S 0$ means:

1. If $i \in S$, then player i is the *only* possible loser of each of i's options.
2. If $i \in S$ but $j \notin S$, then player j has an option of which i is a possible loser. In addition, this option must not add any new possible losers, except possibly player j himself.

For the purposes of the next section, we shall now define the *sum of games*: Define inductively

$$G + H$$

by

$$(G + H)_i = \{G + H^i\} \cup \{G^i + H\}.$$

The sum of games is understood as follows: playing $G + H$ is the same as playing the games G and H side by side, so that i's options in the game $G+H$ should be to either "move in G" or "move in H." If player i chooses to move in G, he chooses an option G^i of the game G, and the game progresses to the position $G^i + H$. Similarly, moving in H moves the game to some position $G+H^i$. Thus, the set of i's options is defined to be the set $\{G+H^i\}\cup\{G^i+H\}$.

3 Number Games

We continue to assume that T is finite, of cardinality n. We shall work with T-tuples of real numbers (or more generally T-tuples of elements of any ordered field F)

$$g = (g_i)_{i \in T} \tag{1}$$

which satisfy

$$\sum_{i \in T} g_i = 0.$$

Obviously, the set of all such T-tuples is an $n-1$-dimensional vector space over F, which we shall denote by F_T. For $S \subseteq T$, and for the T-tuple (1), we now write

$$g <_S 0 \tag{2}$$

for the unique set S of all $i \in T$ with

$$g_i = \min_{k \in T} g_k.$$

Note that S is always non-empty. We also write

$$g \leq_S 0$$

if $g <_U 0$ for some $U \supseteq S$. Note that $g <_T 0$ is equivalent to $g \leq_T 0$ which is equivalent to $g \sim 0$. We shall write

$$g <_S h$$

if $g - h <_S 0$, and similarly for $\leq_S$. By abuse of notation, we write $<_i$ instead of $<_{\{i\}}$.

Lemma 3.1. *If $g <_i h$, then $g_i - g_j < h_i - h_j$ for all $j \neq i$.*

Proof. $g <_i h$ means $g - h <_i 0$, i.e. $g_i - h_i < g_j - h_j$ for all $j \neq i$. □

Below, we shall need the following construction. For $i \in T$, consider the function

$$p_i : F_T \to F_{T-\{i\}}$$

given by

$$p_i(g) = \left(g_j + \frac{g_i}{n-1}\right)_{j \in T-\{i\}}.$$

(In some cases, we shall also use the symbol $p_i g$ instead of $p_i(g)$.)

The function p_i takes n-tuples in F_T and creates $n-1$-tuples in $F_{T-\{i\}}$ in the most natural way: it evenly divides up the strength of the i^{th} element among all the others.

We now proceed to number games. We begin by recalling briefly Conway number games of 2 players [Con01]. The main point is that to each pair of subsets

$$\langle A|B\rangle$$

of the Conway field $\mathcal{C}$, such that for all $a \in A$, $b \in B$ we have

$$a < b,$$

there is assigned an element

$$x = v\langle A|B\rangle \in \mathcal{C} \tag{3}$$

such that, for all $a \in A$, $b \in B$,

$$a < v\langle A|B\rangle < b.$$

More precisely, the Conway field can be defined inductively in this way. Once again, as in the case of games, for every ordinal α there is the "α'th generation" of elements of the Conway field; the 0'th generation consists of the number 0. The Conway field is linearly ordered, and for an element of the form (3) of generation α, A, B are considered its *defining sets* (such pair of sets is required to exist for x to be of the given generation) if all elements of the sets A, B are of generation $< \alpha$. One then refers to elements of A (resp. B) as x_L (resp. x_R), and writes

$$x = \langle x_L|x_R\rangle.$$

One defines addition in the Conway field inductively by

$$x + y = \langle x_L + y, x + y_L | x_R + y, x + y_R\rangle.$$

Now elements represented as (3) in different ways may however be equal. Inductively, an element x of a generation α is ≥ 0 (resp. ≤ 0) if there is no

x_R which is ≤ 0 (resp. x_L which is ≥ 0). One puts $x = 0$ if $0 \leq x \leq 0$. One defines inductively

$$-x = \langle -x_R | -x_L \rangle$$

and $x = y$ if $x + (-y) = 0$. Multiplication is then defined by

$$xy = \langle x_L y + x y_L - x_L y_L, x_R y + x y_R - x_R y_R | x_L y + x y_R - x_L y_R, x_R y + x y_L - x_R y_L \rangle$$

(based on $(x - x_L)(y - y_L) > 0$ etc.).

One must prove that this indeed works. We refer the reader to [Con01] for details, but the following two properties are crucial for our purposes (they follow quite directly from the inductive definition outlined above):

1. $v(\emptyset) = 0$ and $v(G + H) = v(G) + v(H)$ where one defines
$$\langle A|B\rangle + \langle C|D\rangle = \langle v\langle A|B\rangle + C | v\langle C|D\rangle + B\rangle.$$
2. If $C \supseteq A$ and $D \supseteq B$ and for each $x \in C$ (resp. $y \in D$) $x < v\langle A|B\rangle$ (resp. $v\langle A|B\rangle < y$) then
$$v\langle C|D\rangle = v\langle A|B\rangle.$$

Using this, we define inductively a *number game of T players* as a game G for which there exists an n-tuple

$$v(G) \in \mathcal{C}_T$$

such that

1. For all $i \in T$, all G^i are number games, and $v(G^i) <_i v(G)$.
2. For all $i \neq j \in T$,
$$v_i(G) - v_j(G) = v\langle \{v_i(G^i) - v_j(G^i) : p_i v(G^i) \leq_j p_i v(G)\} | \{v_i(G^j) - v_j(G^j) : p_j v(G^j) \leq_i p_j v(G)\}\rangle$$

On first glance, the conditions given in the definition of a number game do not seem natural, but intuitive meaning can be given to them. First, the n-tuple $v(G) = (v_1, v_2, \ldots, v_n)$ gives the strengths of the positions of each player. Larger, positive values of v_i indicate better positions for player i; smaller, negative values indicate worse positions.

Thus, the first statement, that $v(G^i) <_i v(G)$, can be understood as follows: Player i's move from G to G^i not only hurts player i's position; it hurts player i's position more than anyone else's position. This seems natural, as moving in a number game should never "improve" one's position compared to any other player.

The second statement defines the quantity $v_i - v_j$ for each i and j, which is understood to be *i's advantage over j in the game G*. This advantage is defined as the number $v\langle A|B\rangle$, where A is a set of possible advantages i could have after moving, and B is the set of possible advantages j could have after moving. This means that i's advantage in G is more than any advantage he would have after choosing one of his own options G^i, but less than the advantage he would gain were his opponent to move to any G^j.

This would completely explain the definition, however, the sets A and B have an additional restriction on them. Take, for example, the definition of the set A, which contains a condition further restricting its members:

$$A = \{v_i(G^i) - v_j(G^i) : p_i v(G^i) \leq_j p_i v(G)\}.$$

The condition $v(G^i) \leq_j v(G)$ (disregarding the p_i's) would mean that the move from G to G^i must hurt player j's position the most. The $\leq$ allows for this to be nonstrict, namely, that the move may hurt other players just as much. However, in a number game, i's moves from G to G^i must hurt player i the most, so the condition would never be true without considering the p_i.

Recall that the function p_i takes n-tuples and creates $n-1$-tuples, with player i's strength equally distributed among all other players. So, the restriction $p_i v(G^i) \leq_j p_i v(G)$ means that the inequality is true once player i is no longer considered, namely, that the move from G to G^i must hurt player j at least as much as everyone else, with player i himself excluded. Such moves G^i are called *i's anti-j options*, since they do as much damage to player j as possible.

So, number games can be understood as games G for which each player has a well-defined strength of position, given by the n-tuple $v(G)$. G having a *well-defined* strength means that:

1. In the game G, each player's options must be number games, and a player's move must always damage his own position the most.
2. In the game G, *i's advantage over j* is the Conway field element $\langle A|B\rangle$, where A is the set of all advantages player i could have if he chose an *anti-j move*, and B is the set of all advantages player i could have if his opponent chose an *anti-i move*. Thus, player i's advantage over j only depends on the i- and j-options that are primarily directed against one another.

Lemma 3.2. *The T-tuple $v(G)$, if it exists, is uniquely determined.*

Proof. By (1) of the definition of number game and Lemma 2.1, for all G^i we have

$$v_i(G^i) - v_j(G^i) < v_i(G) - v_j(G)$$

while for all G^j we have

$$v_i(G^j) - v_j(G^j) > v_i(G) - v_j(G).$$

By property (2) of Conway games, (2) of the definition of number games then implies

$$v_i(G) - v_j(G) = v\langle v_i(G^i) - v_j(G^i)|v_i(G^j) - v_j(G^j)\rangle \tag{4}$$

which recursively determines $v(G)$. ◻

Lemma 3.3. *A sum of number games is a number game.*

Proof. By induction, both conditions (1), (2) are obviously additive. In particular, in (2), the right hand side for a sum of games contains the Conway sum of the right hand sides of (2) of the individual games, so we can use properties (1) and (2) of Conway games. □

Corollary 3.4. *(of (4))*

$$v(G+H) = v(G) + v(H).$$ □

Proposition 3.5. *If G is a number game and $v(G) <_S 0$, then $G <_S 0$.*

Recall that $v(G) <_S 0$ means that for each $i \in S$, $v_i = \min_{k \in T} v_k$. So, this will show that those players with the least strength of position are exactly those players who will lose a match of this game for some order of play, if all others act in concert.

Proof. Induction. By the induction hypothesis, condition (1) in the definition of number games implies condition (1) for $G <_S 0$.

Suppose condition (2) for number games is valid for G. Choose $i \notin S$, $j \in S$. Then, by definition of $v(G) <_S 0$,

$$v_i(G) > v_j(G).$$

By (2) for number games and properties of Conway games, there is an option G^i such that

$$v_i(G^i) \geq v_j(G^i), \; p_i v(G^i) \leq_j p_i v(G). \tag{5}$$

The second condition implies that

$$v_j(G^i) - v_k(G^i) \leq v_j(G) - v_k(G) \leq 0 \text{ for all } k \neq i, j, \tag{6}$$

so together with (5) this implies that

$$v_j(G^i) = min\{v_p(H) | p \in T\}.$$

On the other hand, if $k \notin S$, (6) implies that

$$v_j(G^i) < v_k(G^i).$$

Thus, $G^i <_T 0$ for some $j \in T \subseteq S \cup \{i\}$, as required in condition 2 for $G <_S 0$. □

Remark. Since for $g \in \mathcal{C}_T$, there is always a unique $S \subseteq T$, $S \neq \emptyset$ with $g <_S 0$, the converse of the Proposition is also true.

Corollary 3.6. *If G is a number game and $v(G) = 0$ then $G \sim 0$.* □

Corollary 3.7. *A number game G has an inverse, i.e. a game H such that $G + H \sim 0$.*

Proof. The symmetric group Σ_T obviously acts on number games by permuting players. Now we obviously have

$$v\left(\sum_{\sigma\in\Sigma_T} \sigma G\right) = \sum_{\sigma\in\Sigma_T} \sigma v(G) = 0,$$

so

$$\sum_{\sigma\in\Sigma_T} \sigma G \sim 0$$

by the previous Corollary. □

4 The Existence Theorem

In this section, we prove that number games of arbitrary values exist.

Theorem 4.1 (Existence theorem). *For every $g \in \mathcal{C}_T$ there exists a number game G with*

$$v(G) = g.$$

Proof. We begin by constructing games that are "worth one move" to each player, then from there games that are "worth x moves" to each player for any $x > 0 \in \mathcal{C}$. Sums and inverses of these games will then be enough to construct a game with $v(G) = g$ for any $g \in \mathcal{C}_T$.

First, we construct the game ${}^{(\mathbf{i})}\mathbf{1}$ for each $i \in T$, the *game worth one move to player i*. It is constructed by

$$\begin{aligned} {}^{(\mathbf{i})}\mathbf{1}_i &= \{0\}, \\ {}^{(\mathbf{i})}\mathbf{1}_j &= \emptyset \text{ for } j \neq i. \end{aligned}$$

In this game, player i has only one option: the move to the zero game. No other players have any options. This is a number game with

$$v({}^{(\mathbf{i})}\mathbf{1}) = {}^{(i)}v \in \mathbb{C}_T$$

where

$${}^{(i)}v_i = \frac{n-1}{n}, \qquad {}^{(i)}v_j = -\frac{1}{n}.$$

For example, for $T = \{1, 2, 3\}$, this construction yields three games:

$$\begin{aligned}
{}^{(1)}\mathbf{1}: &\quad v({}^{(1)}\mathbf{1}) = (\ \ 2/3, -1/3, -1/3)\\
{}^{(2)}\mathbf{1}: &\quad v({}^{(2)}\mathbf{1}) = (-1/3, \ \ 2/3, -1/3)\\
{}^{(3)}\mathbf{1}: &\quad v({}^{(3)}\mathbf{1}) = (-1/3, -1/3, \ \ 2/3)
\end{aligned}$$

To demonstrate how to check that a game is a number game, we will check that indeed the games ${}^{(i)}\mathbf{1}$ above are number games, with $v(G) = ((n-1)/n, -1/n, \ldots, -1/n)$. Obviously, it suffices to consider $i = 1$.

To prove ${}^{(1)}\mathbf{1}$ is a number game, we need to check the two conditions. First, we must check that all the options of G are number games, which they are, and additionally, that $v(G^i) <_i v(G)$ for all i and all G^i.

There is only one option to check, namely $G^i = \mathbf{0}$ and $i = 1$. Indeed, $\mathbf{0}$ is a number game, with $v(\mathbf{0}) = (0,0,\ldots,0)$. So, to check that $v(\mathbf{0}) <_1 v({}^{(1)}\mathbf{1})$, we need only that

$$\begin{aligned}
v(\mathbf{0}) - v({}^{(1)}\mathbf{1}) &<_1 0\\
(0,0,\ldots,0) - ((n-1)/n, -1/n, \ldots, -1/n) &<_1 0\\
(-(n-1)/n, 1/n, \ldots, 1/n) &<_1 0
\end{aligned}$$

which is true. So, condition (1) for number games is satisfied here.

Now, to check condition (2), we need to make sure that the definition's $v_i - v_j$ match up with what we claimed they were by setting $v(G) = ((n-1)/n, -1/n, \ldots, -1/n)$.

First, check $v_1 - v_2$. We should get that $v_1 - v_2 = (n-1)/n + 1/n = 1$. Indeed,

$$\begin{aligned}
v_1 - v_2 &= v\langle \{v_1(G^1) - v_2(G^1) : p_1 v(G^1) \leq_2 p_1 v({}^{(1)}\mathbf{1})\}\\
&\qquad |\{v_1(G^2) - v_2(G^2) : p_2 v(G^2) \leq_1 p_2 v({}^{(1)}\mathbf{1})\}\rangle\\
&= v\langle \{v_1(G^1) - v_2(G^1) : p_1 v(G^1) \leq_2 p_1 v({}^{(1)}\mathbf{1})\} | \emptyset\rangle\\
&= v\langle \{v_1(\mathbf{0}) - v_2(\mathbf{0}) : p_1 v(\mathbf{0}) \leq_2 p_1 v({}^{(1)}\mathbf{1})\} | \emptyset\rangle\\
&= v\langle 0 - 0 | \emptyset\rangle \qquad (\text{we have } (0,\ldots,0) = p_1 v(\mathbf{0}) \leq_2 p_1 v({}^{(1)}\mathbf{1}) = (0,\ldots,0))\\
&= v\langle 0 | \emptyset\rangle\\
&= 1.
\end{aligned}$$

Checking $v_1 - v_i$ is similar for other i. And, checking $v_i - v_j$ for $i, j \neq 1$ ($i \neq j$) is easy, since we want $v_i - v_j = 0$, and indeed, it is

$$\begin{aligned}
&v\langle \{v_i(G^i) - v_j(G^i) : p_i v(G^i) \leq_j p_i v({}^{(1)}\mathbf{1})\} |\\
&\quad \{v_i(G^j) - v_j(G^j) : p_j v(G^j) \leq_i p_j v({}^{(1)}\mathbf{1})\}\rangle
\end{aligned}$$

$$= v\langle\emptyset|\emptyset\rangle$$
$$= 0.$$

Now, we construct games ${}^{(\mathbf{i})}\mathbf{x}$ for all numbers $x \in \mathbb{C}$. The game ${}^{(\mathbf{i})}\mathbf{x}$ is *the game worth x moves to player i.*

Lemma 4.2. *Given $x \geq 0 \in \mathbb{C}$ and $i \in T$, there exists a number game ${}^{(\mathbf{i})}\mathbf{x}$ such that*

$$v({}^{(\mathbf{i})}\mathbf{x}) = x \cdot v({}^{(\mathbf{i})}\mathbf{1}).$$

For $x = 0$, the zero game satisfies this condition, so we need only consider $x > 0$. Once this construction is complete, the proof will be nearly finished.

Given $x \in \mathbb{C}$, we know that it is constructed by

$$x = v\langle L|R\rangle$$

for some sets L, R of simpler numbers in $\mathcal{C}$. In addition, since $x > 0$, we have that all $x^R > 0$. We can assume without loss of generality that all $x^L \geq 0$ as well, so by induction, we may assume that we have already constructed the "simpler" number games ${}^{(\mathbf{j})}\mathbf{x}^{\mathbf{L}}$ and ${}^{(\mathbf{j})}\mathbf{x}^{\mathbf{R}}$ for all $x^L \in L$, $x^R \in R$, and $j \in T$.

Now define an inverse ${}^{(\mathbf{j})}-\mathbf{x}^{\mathbf{R}}$ of ${}^{(\mathbf{j})}\mathbf{x}^{\mathbf{R}}$ as the sum of previously constructed number games:

$${}^{(\mathbf{j})}-\mathbf{x}^{\mathbf{R}} = \sum_{k \neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}.$$

To prove that ${}^{(\mathbf{j})}-\mathbf{x}^{\mathbf{R}}$ is an inverse of ${}^{(\mathbf{j})}\mathbf{x}^{\mathbf{R}}$, note that by Corollary 3.7, we have that every game G has an inverse given by the sum of all the games that are the result of permuting the players of G. So, it suffices to show that ${}^{(\mathbf{j})}-\mathbf{x}^{\mathbf{R}}$ is the sum of the $(n!-1)$ number games which are permutations of the game G. These permutations are still number games, and they are all given by ${}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}$ for some k:

$$\begin{aligned}
{}^{(\mathbf{j})}-\mathbf{x}^{\mathbf{R}} &= \left(((n-1)!-1)\cdot {}^{(\mathbf{j})}\mathbf{x}^{\mathbf{R}}\right) + \left(((n-1)!)\cdot \sum_{k\neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}\right)\\
&= \left(((n-1)!-1)\cdot {}^{(\mathbf{j})}\mathbf{x}^{\mathbf{R}}\right) + \left(((n-1)!-1)\cdot \sum_{k\neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}\right) + \left(\sum_{k\neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}\right)\\
&= \left(((n-1)!-1)\cdot \sum_{\text{all k}} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}\right) + \left(\sum_{k\neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}\right)\\
&= \left(((n-1)!-1)\cdot \mathbf{0}\right) + \left(\sum_{k\neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}\right)\\
&= \sum_{k\neq j} {}^{(\mathbf{k})}\mathbf{x}^{\mathbf{R}}.
\end{aligned}$$

Now consider the game G given by

$$G_i = \{{}^{(\mathrm{i})}\mathbf{x}^{\mathbf{L}}\},$$
$$G_j = \{{}^{(\mathrm{j})}-\mathbf{x}^{\mathbf{R}}\}, \text{ for } j \neq i.$$

Claim. G is the game ${}^{(\mathrm{i})}\mathbf{x}$ that satisfies the lemma, i.e. it is a number game with $v(G) = x \cdot v({}^{(\mathrm{i})}\mathbf{1})$.

To show that G is a number game with the given tuple, we need to check the conditions (1) and (2) for a number game.

For condition (1), we first need that each option of G is a number game, which is true by induction. Then we must show that each option G^k has $v(G^k) <_k v(G)$.

For player i, then, we must show that

$$v({}^{(\mathrm{i})}\mathbf{x}^{\mathbf{L}}) <_i x \cdot v({}^{(\mathrm{i})}\mathbf{1}).$$

But by induction, the left-hand side is given by

$$x^L \cdot v({}^{(\mathrm{i})}\mathbf{1}),$$

so it must be proven that

$$x^L \cdot v({}^{(\mathrm{i})}\mathbf{1}) <_i x \cdot v({}^{(\mathrm{i})}\mathbf{1})$$
$$(x^L - x) \cdot v({}^{(\mathrm{i})}\mathbf{1}) <_i 0.$$

This is true: since $x^L - x$ is negative, while the only positive entry of $v({}^{(\mathrm{i})}\mathbf{1})$ is in the i^{th} position, we have that the only negative value of the n-tuple is in the i^{th} position. So, it is $<_i 0$.

For other players j, we must show that

$$v({}^{(\mathrm{j})} - \mathbf{x}^{\mathbf{R}}) <_j x \cdot v({}^{(\mathrm{i})}\mathbf{1}).$$

But by induction, we already know $v({}^{(\mathrm{j})} - \mathbf{x}^{\mathbf{R}})$ is a number game, and since it is the inverse of ${}^{(\mathrm{j})}\mathbf{x}^{\mathbf{R}}$, we know $v({}^{(\mathrm{j})} - \mathbf{x}^{\mathbf{R}}) = -v({}^{(\mathrm{j})}\mathbf{x}^{\mathbf{R}})$. So, we need to show

$$v({}^{(\mathrm{j})} - \mathbf{x}^{\mathbf{R}}) <_j x \cdot v({}^{(\mathrm{i})}\mathbf{1})$$
$$-v({}^{(\mathrm{j})}\mathbf{x}^{\mathbf{R}}) <_j x \cdot v({}^{(\mathrm{i})}\mathbf{1})$$
$$-x^R \cdot v({}^{(\mathrm{j})}\mathbf{1}) <_j x \cdot v({}^{(\mathrm{i})}\mathbf{1})$$
$$-x^R \cdot v({}^{(\mathrm{j})}\mathbf{1}) - x \cdot v({}^{(\mathrm{i})}\mathbf{1}) <_j 0.$$

The n-tuple on the left-hand side has positive entries for every index other than i and j, but negative entries for indices i and j. However, the j^{th} entry is most negative, since $x^R > x$, so the inequality holds.

This finishes the verification of the first condition for G to be a number game with $v(G)$ as desired.

To finish the proof that G is a number game, we must show condition (2), that for each $j, k \in T$, the difference $v_j(G) - v_k(G)$ is the same as what is required. We split this into two cases: one where both $j, k \neq i$, and one where one of them does equal i.

1. Check condition (2) for $v_j - v_k$, where $j, k \neq i$.
   We have claimed that G is a number game with $v(G) = x \cdot v({}^{(i)}\mathbf{1})$, so we claim that $v_j - v_k = (x \cdot v({}^{(i)}\mathbf{1}))_j - (x \cdot v({}^{(i)}\mathbf{1}))_k$. This equals zero since $v({}^{(i)}\mathbf{1})_j = v({}^{(i)}\mathbf{1})_k$ for $j, k \neq i$.
   So, we must show that the right-hand side of the equation in condition (2) is $0 \in \mathbb{C}$. The right-hand side of the equation reads as follows:
   $$v\langle \{v_j(G^j) - v_k(G^j) : p_j v(G^j) \leq_k p_j v(G)\} | \\ \{v_j(G^k) - v_k(G^k) : p_k v(G^k) \leq_j p_k v(G)\}\rangle.$$
   Now, a number $y = v\langle L|R\rangle$ is $0 \in \mathbb{C}$ if and only if all $y^L < 0$ and all $y^R > 0$ (if they exist). We shall thus show that all the left options, if they exist, are less than zero, and all the right options, if they exist, are greater than zero:
   $$\begin{aligned} v_j(G^j) - v_k(G^j) < 0 &\quad \text{when } p_j v(G^j) \leq_k p_j v(G), \\ v_j(G^k) - v_k(G^k) > 0 &\quad \text{when } p_k v(G^k) \leq_j p_k v(G), \\ \text{which} &\quad \text{simplifies to} \\ v_j(G^j) < v_k(G^j) &\quad \text{when } p_j v(G^j) \leq_k p_j v(G), \\ v_j(G^k) > v_k(G^k) &\quad \text{when } p_k v(G^k) \leq_j p_k v(G). \end{aligned}$$
   In fact, we shall show more strongly that $v_j(G^j) < v_k(G^j)$ and $v_j(G^k) > v_k(G^k)$ always, regardless of the restrictions. Additionally, these two statements are symmetric in j and k, so we need only show the first one.
   Since all the G^j are of the form ${}^{(j)}{-\mathbf{x}^{\mathbf{R}}}$, we only need to show that
   $$v_j({}^{(j)}{-\mathbf{x}^{\mathbf{R}}}) < v_k({}^{(j)}{-\mathbf{x}^{\mathbf{R}}}).$$
   But this is true, since $x^R > 0$ tells us that
   $$\begin{aligned} v_j({}^{(j)}{-\mathbf{x}^{\mathbf{R}}}) &= (-x^R) \cdot (n-1)/n < 0, \\ v_k({}^{(j)}{-\mathbf{x}^{\mathbf{R}}}) &= (-x^R) \cdot (-1/n) > 0. \end{aligned}$$
   $$\text{So } v_j({}^{(j)}{-\mathbf{x}^{\mathbf{R}}}) < 0 < v_k({}^{(j)}{-\mathbf{x}^{\mathbf{R}}}).$$

   This shows that $v_j - v_k = 0$, as desired.

2. Check condition (2) for $v_i - v_j$, where $j \neq i$.

   We have claimed that G is a number game with $v(G) = x \cdot v({}^{(\mathbf{i})}\mathbf{1})$. So, we have claimed that $v_i - v_j$ is as follows:

$$\begin{aligned} v_i - v_j &= x \cdot (v_i({}^{(\mathbf{i})}\mathbf{1}) - v_j({}^{(\mathbf{i})}\mathbf{1})) \\ &= x \cdot (\frac{n-1}{n} - \frac{-1}{n}) \\ &= x. \end{aligned}$$

   So, we must show that the right-hand side of condition (2) for number games really yields the number $x \in \mathbb{C}$. The right-hand side is here the number $v\langle A|B\rangle$, where

$$\begin{aligned} v\langle A|B\rangle = v\langle\, &\{v_i(G^i) - v_j(G^i) : p_i v(G^i) \leq_j p_i v(G)\}| \\ &\{v_i(G^j) - v_j(G^j) : p_j v(G^j) \leq_i p_j v(G)\}\rangle. \end{aligned}$$

   First, look at the set $A = \{v_i(G^i) - v_j(G^i) : p_i v(G^i) \leq_j p_i v(G)\}$. Using the fact that we know all the G^i and G^j yields

$$\begin{aligned} A &= \{v_i(G^i) - v_j(G^i) : p_i v(G^i) \leq_j p_i v(G)\} \\ &= \{v_i({}^{(\mathbf{i})}\mathbf{x}^{\mathbf{L}}) - v_j({}^{(\mathbf{i})}\mathbf{x}^{\mathbf{L}}) : p_i v({}^{(\mathbf{i})}\mathbf{x}^{\mathbf{L}}) \leq_j p_i v(G)\} \\ &= \{\big(x^L \cdot ((n-1)/n)\big) - \big(x^L \cdot (-1/n)\big) : p_i v({}^{(\mathbf{i})}\mathbf{x}^{\mathbf{L}}) \leq_j p_i v(G)\} \\ &= \{x^L : p_i v({}^{(\mathbf{i})}\mathbf{x}^{\mathbf{L}}) \leq_j p_i v(G)\} \end{aligned}$$

   So, $A \subseteq L$. Deciphering the condition at right yields that

$$\begin{aligned} A &= \{x^L : p_i v({}^{(\mathbf{i})}\mathbf{x}^{\mathbf{L}}) \leq_j p_i v(G)\} \\ &= \{x^L : p_i(x^L \cdot \overset{(i)}{v}) \leq_j p_i(x \cdot \overset{(i)}{v})\}. \\ &= \{x^L : p_i(x^L \cdot \overset{(i)}{v}) - p_i(x \cdot \overset{(i)}{v}) \leq_j 0\}. \end{aligned}$$

   We will show that $A = L$, by showing that the right-hand condition is actually true for all x^L.

   Consider the j^{th} element of the n-tuple in the condition:

$$\big(p_i(x^L \cdot \overset{(i)}{v}) - p_i(x \cdot \overset{(i)}{v})\big)_j.$$

$$\begin{aligned} \big(p_i(x^L \cdot \overset{(i)}{v}) - p_i(x \cdot \overset{(i)}{v})\big)_j &= \big(p_i(x^L \cdot \overset{(i)}{v})\big)_j - \big(p_i(x \cdot \overset{(i)}{v})\big)_j \\ &= x^L \cdot \big(\overset{(i)}{v}_j + \frac{1}{n-1}\overset{(i)}{v}_i\big) - x \cdot \big(\overset{(i)}{v}_j + \frac{1}{n-1}\overset{(i)}{v}_i\big) \end{aligned}$$

$$
\begin{aligned}
&= (x^L - x) \cdot \big(\overset{(i)}{v}_j + \frac{1}{n-1} \overset{(i)}{v}_i \big) \\
&= (x^L - x) \cdot \big(-\frac{1}{n} + \frac{1}{n-1} \cdot \frac{n-1}{n} \big) \\
&= (x^L - x) \cdot 0 \\
&= 0.
\end{aligned}
$$

Thus $p_i(x^L \cdot \overset{(i)}{v}) - p_i(x \cdot \overset{(i)}{v}) = \mathbf{0}$. So, it is $\leq_j 0$ for all $j \neq i$. Thus $A = \{\text{all } x^L\} = L$.

Now consider the set $B = \{v_i(G^j) - v_j(G^j) : p_j v(G^j) \leq_i p_j v(G)\}$. Using the fact that we know all the G^i and G^j yields, analogously to the case of A,

$$
\begin{aligned}
B &= \{v_i(G^j) - v_j(G^j) : p_j v(G^j) \leq_i p_j v(G)\} \\
&= \{v_i(\overset{(\mathbf{j})}{-\mathbf{x}^{\mathbf{R}}}) - v_j(\overset{(\mathbf{j})}{-\mathbf{x}^{\mathbf{R}}}) : p_j v(\overset{(\mathbf{j})}{-\mathbf{x}^{\mathbf{R}}}) \leq_i p_j v(G)\} \\
&= \{\big(-x^R \cdot (-1/n)\big) - \big(-x^R \cdot ((n-1)/n)\big) : p_j v(\overset{(\mathbf{j})}{-\mathbf{x}^{\mathbf{R}}}) \leq_i p_j v(G)\} \\
&= \{x^R : p_j v(\overset{(\mathbf{j})}{-\mathbf{x}^{\mathbf{R}}}) \leq_i p_j v(G)\}
\end{aligned}
$$

So, $B \subseteq R$. Similarly as before, we will show $B = R$. So, looking closely at the condition at right leads to

$$
\begin{aligned}
B &= \{x^R : p_j v(\overset{(\mathbf{j})}{-\mathbf{x}^{\mathbf{R}}}) \leq_i p_j v(G)\} \\
&= \{x^R : p_j(-x^R \cdot \overset{(j)}{v}) \leq_i p_j(x \cdot \overset{(j)}{v})\} \\
&= \{x^R : p_j(-x^R \cdot \overset{(j)}{v}) - p_j(x \cdot \overset{(j)}{v}) \leq_i 0\}.
\end{aligned}
$$

Again the final step is to show that $p_j(-x^R \cdot \overset{(j)}{v}) - p_j(x \cdot \overset{(j)}{v}) = \mathbf{0}$, so that it is in particular $\leq_i 0$.

So consider $\big(p_j(-x^R \cdot \overset{(j)}{v}) - p_j(x \cdot \overset{(j)}{v})\big)_k$. Again,

$$
\begin{aligned}
\big(p_j(-x^R \cdot \overset{(j)}{v}) - p_j(x \cdot \overset{(j)}{v})\big)_k &= \big(p_j(-x^R \cdot \overset{(j)}{v})\big)_k - \big(p_j(x \cdot \overset{(j)}{v})\big)_k \\
&= -x^R \cdot \big(\overset{(j)}{v}_k + \frac{1}{n-1} \overset{(j)}{v}_j \big) - x \cdot \big(\overset{(j)}{v}_k + \frac{1}{n-1} \overset{(j)}{v}_j \big) \\
&= (-x^R - x) \cdot \big(\overset{(j)}{v}_k + \frac{1}{n-1} \overset{(j)}{v}_j \big) \\
&= (-x^R - x) \cdot \big(-\frac{1}{n} + \frac{1}{n-1} \cdot \frac{n-1}{n} \big) \\
&= (-x^R - x) \cdot 0 \\
&= 0.
\end{aligned}
$$

Thus the condition was true for all x^R, and $B = \{\text{all } x^R\} = R$.

So, we have $A = L$, $B = R$, which implies that, as desired,

$$x = v\langle L|R\rangle = v\langle A|B\rangle.$$

This concludes the proof of the lemma, that number games of the form $^{(\mathbf{i})}\mathbf{x}$ can be constructed for all $x \in \mathbb{C}$, $i \in T$.

Now that these games have been constructed, the proof of the theorem can be completed: given a T-tuple $(v_1, v_2, \ldots, v_n) \in \mathbb{C}_T$, it is a linear combination of the T-tuples $\overset{(1)}{v}, \overset{(2)}{v}, \ldots, \overset{(n-1)}{v}$, of the form

$$(v_1, v_2, \ldots, v_n) = \sum_{i=1}^{n-1} a_i \cdot \overset{(i)}{v}$$

for some "scalars" $a_i \in \mathbb{C}$.

Then a number game G with $v(G) = (v_1, v_2, \ldots, v_n)$ is constructed by

$$G = \sum_{i=1}^{n-1} {}^{(\mathbf{i})}\mathbf{a_i}.$$

□

Appendix 1: Examples of Diagrams of Three-Player Games

By the results of the previous section, we know that number games of any position strength can be constructed, and that for games of n players, the games form an $n-1$-dimensional vector space over the Conway field $\mathcal{C}$. In [Con01], one often uses the "number line" to visualize the position strengths of games of two players. When three player games are considered, we have a 2-dimensional vector space, i.e. a "number plane." To help understand the meaning of the results of this paper, this appendix is dedicated to the development of a similar visualization of the "number plane" of games of three players.

To begin, we place the zero game $\mathbf{0}$ at the origin. The next games to be constructed, $^{(\mathbf{1})}\mathbf{1}$, $^{(\mathbf{2})}\mathbf{1}$, and $^{(\mathbf{3})}\mathbf{1}$, are placed next. These are the games worth one move to each of players 1, 2, and 3 respectively, so we place these games on the plane one unit from the origin. Let us separate them by 120° angles, with $^{(\mathbf{1})}\mathbf{1}$ at the lower right, $^{(\mathbf{2})}\mathbf{1}$ at the lower left and $^{(\mathbf{3})}\mathbf{1}$ at the top as follows:

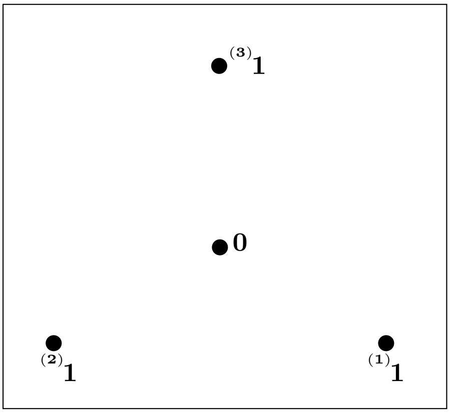


Fig. 1. The beginning of the number plane

At this point, it becomes clear what must be the location of each of the number games. All games of the form ${}^{(1)}\mathbf{x}$ for $x > 0$ must lie on a ray through the zero game and ${}^{(1)}\mathbf{1}$; similarly for the other players. At that point, all other games are constructed by sums of those games, and are placed on the plane by vector addition.

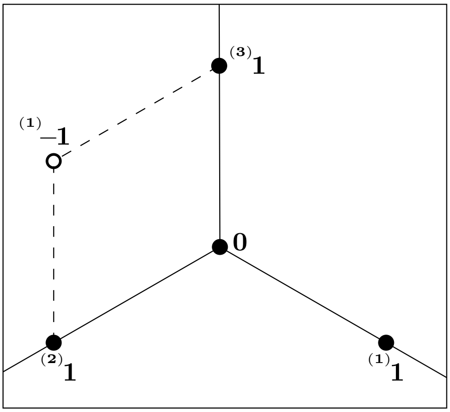


Fig. 2. The placement of ${}^{(1)}-\mathbf{1} = {}^{(2)}\mathbf{1} + {}^{(3)}\mathbf{1}$

In this way, the plane is partitioned into three open regions, consisting of the games which are $<_1 0$, $<_2 0$, and $<_3 0$. These games will be lost by player 1, 2, and 3 respectively, if the others work in concert.

For example, all the games $<_1 0$ are in the region of the plane opposite the game ${}^{(1)}\mathbf{1}$, which is worth one move to player 1. The game ${}^{(1)}\mathbf{1}$ itself lies on the boundary between the regions $<_2 0$ and $<_3 0$, since it is itself $<_{\{2,3\}} 0$: the loser will depend on the order of play.

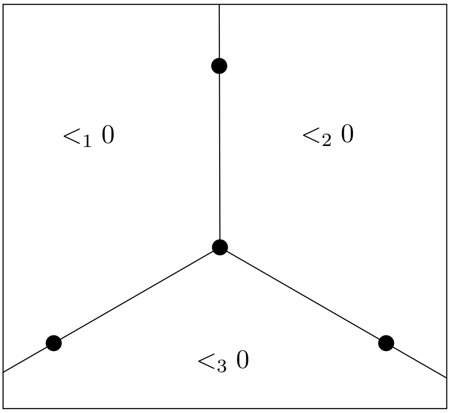


Fig. 3. Partitioning the number plane by who loses the game

While the preceding partition of the plane is very natural, we need not restrict ourselves to comparing number games to **0**. For example, we may partition the plane by comparing the games to our first constructed game, ${}^{(1)}\mathbf{1}$. Thus, the region i consists of all the games "worse than ${}^{(1)}\mathbf{1}$ for player i."

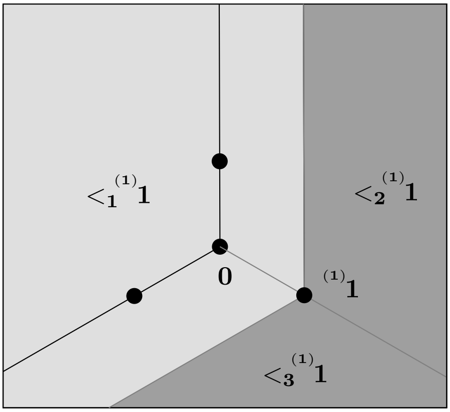


Fig. 4. Comparing games to ${}^{(1)}\mathbf{1}$

This visualization provides a means of testing the first condition for a game to be a number game. The condition was:

$$\text{For all } i \in T, \text{ all } G^i \text{ are number games, and } v(G^i) <_i v(G).$$

For example, suppose that we were to test the claim that the game

$$G = ({}^{(1)}-\mathbf{1}, {}^{(2)}-\mathbf{1}, \mathbf{0}),$$

which has one option for each player, is really deserving of the name ${}^{(3)}(\mathbf{1/2})$, the game worth 1/2 move to the third player. Then we only need to plot the positions of the options, and the position of the "expected" value:

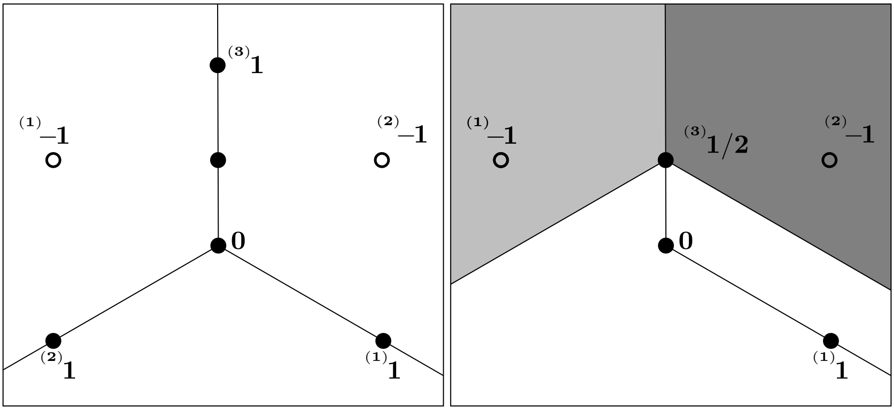


Fig. 5. At left, the options are plotted on the plane. At right, they are shown compared to the expected value ${}^{(3)}(\mathbf{1/2})$

The figure above shows that player 1's option in G, which is the game ${}^{(1)}-\mathbf{1}$, is in the region $<_1 {}^{(3)}(\mathbf{1/2})$. Similarly, we have ${}^{(2)}-\mathbf{1} <_2 {}^{(3)}(\mathbf{1/2})$ and $\mathbf{0} <_3 {}^{(3)}(\mathbf{1/2})$. This visually checks the first condition for the game G to actually be the game ${}^{(3)}(\mathbf{1/2})$.

The number plane can also help in visualizing the second condition for a game to be a number game. In the second condition, we need that the difference $v_i - v_j$ in the tuple $v(G)$ is given by a particular number; that number is defined by the options of players i and j.

As previously discussed, the number $v_i - v_j$ must depend only on player i's *anti-j options*, and player j's *anti-i options*. These can be given graphical meaning as well.

One of i's *anti-j options* G^i of G must first be one of i's options in G, so we first must have $G^i <_i G$. However, the additional restriction $p_i(G^i) \leq_j p_i(G)$ means that disregarding player i, player j must be hurt most by the move.

Thus, an *anti-j option* of G must lie in the region $<_i G$, but it must also be closer to the $<_j G$ region than any $<_k G$ region. The plane then looks like this:

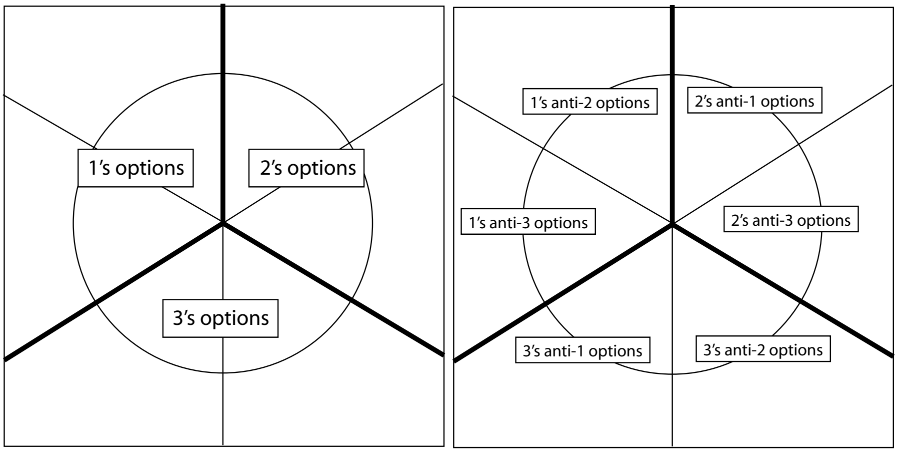


Fig. 6. Finding whose options are anti-who

Returning to the case of the game ${}^{(\mathbf{3})}(\mathbf{1/2})$, we can use this reference to classify each option in the game.

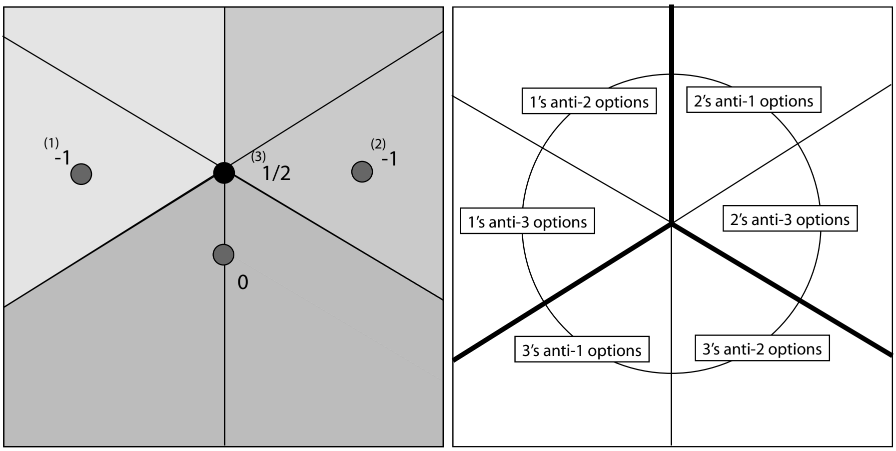


Fig. 7. Classifying the options of ${}^{(\mathbf{3})}(\mathbf{1/2})$

So, player 1's option ${}^{(\mathbf{1})}-\mathbf{1}$ is an *anti-3* option,
Player 2's option ${}^{(\mathbf{2})}-\mathbf{1}$ is an *anti-3* option,
and player 3's option $\mathbf{0}$ is both *anti-1* and *anti-2*.

So, we calculate $v_1 - v_2$ to be 0, since player 1 has no anti-2 options, and player 2 has no $anti-1$ options.

$v_1 - v_3$ must be given by

$$\begin{aligned} v_1 - v_3 &= \left\langle v_1\left({}^{(1)}-\mathbf{1}\right) - v_3\left({}^{(1)}-\mathbf{1}\right) \middle| v_1(\mathbf{0}) - v_3(\mathbf{0})\right\rangle \\ &= \langle(-2/3) - (1/3)|0 - 0\rangle = \langle -1|0\rangle \\ &= -1/2, \text{ as desired. Player 3 has a 1/2-move advantage.} \end{aligned}$$

$v_2 - v_3$ would be calculated similarly.

Appendix 2: Why Do We Define Number Games in This Way?

At first glance, our definition of a number game of multiple players may seem somewhat artificial. In particular, it is not obvious why we need to include in condition (2) the restriction on options which can be considered during evaluation of the $(n-1)$-tuple associated with the n-player game. To determine the advantage player i has over player j, we only use i's *anti-j options* and j's *anti-i options*. It seems that it would be more natural to include all the options on both sides.

However, it turns out that allowing all options to be considered would cause our theory of number games to break down. To be more precise, let us, for the moment, provisionally define a *quasi number game of T players* as a game G for which there exists an n-tuple

$$v(G) \in \mathcal{C}_T$$

such that

1. For all $i \in T$, all G^i are quasi number games, and $v(G^i) <_i v(G)$.
2. For all $i \neq j \in T$,

$$v_i(G) - v_j(G) = v\langle\{v_i(G^i) - v_j(G^i)\}|\{v_i(G^j) - v_j(G^j)\}\rangle$$

The problem with this definition is as follows.

Proposition 4.3. *There exists a quasi number game that does not possess a set of losers as defined in Section 2.*

Proof. Consider the following game:

$$G = \left({}^{(3)}\mathbf{1}, {}^{(1)}\mathbf{1} + {}^{(3)}\mathbf{2}, \mathbf{0}\right)$$

Each player has one option, and all the options are number games. We shall show that G is a quasi number game. First, note that every number game is a quasi number game by the proof of Lemma 3.2. Next, every option of G is a number game, and hence a quasi number game. In fact, we find the 3-tuples associated with each of the options of G to be:

$$\begin{aligned} v(\overset{(3)}{}\mathbf{1}) &= (-1/3, -1/3, 2/3) \\ v(\overset{(1)}{}\mathbf{1} + \overset{(3)}{}\mathbf{2}) &= (0, -1, 1) \\ v(\mathbf{0}) &= (0, 0, 0). \end{aligned}$$

Now we must have for G:

$$\begin{aligned} v_1(G) - v_2(G) &= v\langle\{v_1(\overset{(3)}{}\mathbf{1}) - v_2(\overset{(3)}{}\mathbf{1})\}|\{v_1(\overset{(1)}{}\mathbf{1} + \overset{(3)}{}\mathbf{2}) - v_2(\overset{(1)}{}\mathbf{1} + \overset{(3)}{}\mathbf{2})\}\rangle \\ &= v\langle(-1/3) - (-1/3)|(0) - (-1)\rangle \\ &= v\langle 0|1\rangle \\ &= 1/2 \\ v_1(G) - v_3(G) &= v\langle\{v_1(\overset{(3)}{}\mathbf{1}) - v_3(\overset{(3)}{}\mathbf{1})\}|\{v_1(\mathbf{0}) - v_3(\mathbf{0})\}\rangle \\ &= v\langle(-1/3) - (2/3)|(0) - (0)\rangle \\ &= v\langle -1|0\rangle \\ &= -1/2 \\ v_2(G) - v_3(G) &= v\langle\{v_2(\overset{(1)}{}\mathbf{1} + \overset{(3)}{}\mathbf{2}) - v_3(\overset{(1)}{}\mathbf{1} + \overset{(3)}{}\mathbf{2})\}|\{v_2(\mathbf{0}) - v_3(\mathbf{0})\}\rangle \\ &= v\langle(-1) - (1)|(0) - (0)\rangle \\ &= v\langle -2|0\rangle \\ &= -1 \end{aligned}$$

Then, $v_1(G), v_2(G)$, and $v_3(G)$ must satisfy the equations

$$\begin{aligned} v_1 - v_2 &= 1/2 \\ v_1 - v_3 &= -1/2 \\ v_2 - v_3 &= -1. \end{aligned} \tag{7}$$

The tuple $(0, -1/2, 1/2)$ satisfies these equations, so G is be a quasi number game under the above definition.

However, the game G has no set of possible losers S. To see this, suppose such set S existed.

First, player 1 cannot be a member of S, because under the order of play 1-2-3, player 1 would be the first member of S to move, and so it would have to be possible for player 1 to lose the game G. However, player 2 is the loser in this case.

Second, player 3 cannot be a member of S, because under the order of play 3-2-1, player 3 would be the first member of S to move, and so it would have to be possible for player 3 to lose the game G. However, player 2 is the loser in this case.

Finally, player 2 cannot be a member of S, because if that were true, we would have to have $S = \{\mathbf{2}\}$. However, under the order of play 3-1-2, player 2 is the first member of S to move, and player 1 is still the loser.

Thus, the game G has no set of possible losers S, but is a quasi number game, thus completing the proof of the Proposition. □

For completeness, let us see explicitly why the game G considered in the above proof is not an actual number game. It turns out that the only all of the calculations in the proof remain valid for the number game definition, except restricting to anti-2 and anti-3 options changes the last equation of (7) to

$$v_2(G) - v_3(G) = v\langle\emptyset|\emptyset\rangle = 0.$$

This change will render (7) inconsistent, and thus G is not a number game.

To explain this intuitively, there being no set of possible losers S means roughly that it is sometimes to one's advantage to move first in the game G. In a number game, it should never be to a player's advantage to move first, as a move should always imply a consumption of one's limited resources. The game G does not fit this model, so the relaxed definition must be rejected in favor of the one that we have proposed.

References

[Con01] J. H. Conway: *On numbers and games. Second edition*, A K Peters, Ltd., Natick, MA, 2001. xii+242 pp. ISBN: 1-56881-127-6

[Con72] J. H. Conway: All numbers, great and small, *The Univ. of Calgary Math. Research Paper* no. 149, February, 1972

[Knu74] D. E. Knuth: *Surreal numbers*, Addison-Wesley, 1974

[Now02] R. Nowakowski (ed.): *More games of no chance*, Proceedings of the 2nd Combinatorial Games Theory Workshop held in Berkeley, CA, July 24-28, 2000, Math. Sciences research Institute Publ. 42, Cambridge Univ. Press, Cambridge, 2002

Two Isoperimetric Problems for Euclidean Simplices

Miroslav Fiedler*

Academy of Sciences of the Czech Republic, Institute of Computer Science,
Pod vodárenskou věží 2, 182 07 Praha 8, Czech Republic
fiedler@math.cas.cz

Summary. Best possible estimates for the lengths of a Hamiltonian path and of a Hamiltonian polygon on a Euclidean simplex of given volume are given. The extreme cases are described.

AMS Subject Classification. 52A40, 52B11.
Keywords. Simplex, Hamiltonian path, Hamiltonian polygon, isoperimetric problem.

1 Introduction

Let Σ be a simplex in a Euclidean space. Using graph theory terminology, we call *Hamiltonian path on* Σ every open polygon which consists of edges of Σ and contains all the vertices. Also, a *Hamiltonian polygon* on Σ is every closed polygon consisting of edges and containing all the vertices. It is immediate that for a simplex of given volume there is a bound from below for the lengths of both these polygons. We intend to give the best possible bounds and completely describe the cases in which equality is attained.

In [Fie61], we proved the following theorem on so called nets on a simplex. We defined a *net* on a simplex as a subset of the set of all edges, such that every vertex of the simplex belongs to some edge in the net. A net is called *metric* if for each edge of the net its length is given.

Theorem 1.1 ([Fie61], Theorem 3.1). *Let a metric net N on a simplex be given. There exists a simplex of maximum volume with this net if and only if the net N is connected, i. e. if it is possible to pass from any vertex of the net to any other vertex using edges in the net only. In addition, every simplex with this maximum volume has the property that every interior angle opposite an edge of the simplex not belonging to N is right.*

* Research supported by grant A1030302.

We intend to apply this theorem to two most important cases; namely, that the net is a path, and, that the net is a (closed) polygon. For simplicity, we call an edge in the net N a *specified edge*, an edge not belonging to N an *unspecified edge.*

2 Results

Theorem 2.1. *The n-dimensional volume of an n-simplex with vertices A_1, ..., A_{n+1}, for which the path $A_1 A_2 \cdots A_{n+1}$ has length ℓ, does not exceed $\frac{1}{n!}[\frac{\ell}{n}]^n$.*

The equality is attained if and only if the vertices A_i can be completed to 2^n vertices of a cube in such a way that all edges $A_i A_{i+1}$ ($i = 1, \ldots, n+1$, $A_{n+2} \equiv A_1$) have length $\frac{\ell}{n}$ and are mutually perpendicular. The simplex is called the Schlaefli simplex.

Proof. By Theorem 1.1 the maximum volume of such a simplex exists. We call it Σ. In addition, all interior angles opposite the unspecified edges are right.

Recall (cf. [Fie57]) that every n-simplex has at least n acute interior angles, and there exist n-simplices which have n acute interior angles and all the remaining $\binom{n}{2}$ interior angles right. These simplices were called *right simplices* in [Fie57]. The edges opposite acute angles are called *legs*; they are mutually perpendicular.

By the theory of right simplices ([Fie57]), the simplex Σ is a right simplex and its legs coincide with n mutually perpendicular edges of a box (orthogonal parallelepiped). The volume of such simplex is thus $\frac{1}{n!}$ times the volume of the box which is the product of the lengths of the legs. Since the sum of the lengths of the legs is ℓ, the maximum product is $[\frac{\ell}{n}]^n$. Equality is attained exactly in the case described. □

We call *span* of an n-simplex the length of the shortest Hamiltonian path of the simplex.

Corollary 2.2. *The span of an n-simplex with n-dimensional volume V is always greater than or equal to $n\sqrt[n]{n!V}$.*

Equality is attained if and only if the simplex is the Schlaefli simplex on a cube with volume $n!V$ and the path is the path of its legs.

To investigate Hamiltonian polygons on a simplex, we need to recall some facts about cyclic simplices (cf. [Fie61]). An n-simplex with vertices $A_1, \ldots, A_{n+1}$ is *cyclic* with the distinguished polygon $A_1 A_2 \cdots A_{n+1} A_1$ if all interior angles except those opposite the edges of the polygon are right, and none of the angles opposite the edges of the polygon is right. We also call a cyclic simplex *regularly cyclic* if all edges of the distinguished polygon have equal length.

Theorem 2.3 ([Fie61], Theorem 2.6). *Let $A_1, \dots, A_{n+1}$ be vertices of an n-simplex. Then it is cyclic with distinguished polygon $A_1A_2 \cdots A_{n+1}A_1$ if and only if there exist non-zero numbers $p_1, \dots, p_{n+1}$ such that the inner products of the vectors $\mathbf{a}_i = A_{i+1} - A_i$ $(i = 1, \dots, n+1,\ A_{n+2} \equiv A_1)$ satisfy*

$$\langle \mathbf{a}_i, \mathbf{a}_j \rangle = -\frac{1}{p} p_i p_j,\ i \neq j,$$
$$\langle \mathbf{a}_i, \mathbf{a}_i \rangle = \frac{1}{p} p_i (p - p_i),$$

where $p = \sum_i p_i$.

Remark 2.4. Only two cases are possible.

Case A. All p_i's are positive.
Case B. Exactly one p_i is negative and p is negative as well.

In the first case, the squares ℓ_i^2 of the lengths of the $\mathbf{a}_i$'s satisfy the strict polygonal inequality

$$\ell_i^2 < \sum_{j \neq i} \ell_j^2$$

for each i, which can be written as

$$2 \max_i \ell_i^2 < \sum_i \ell_i^2.$$

In the second case,

$$2 \max_i \ell_i^2 > \sum_i \ell_i^2. \tag{1}$$

We can now formulate the theorem:

Theorem 2.5. *Let ℓ be the length of the closed polygon $A_1A_2 \cdots A_{n+1}A_1$, where $A_1, A_2, \dots, A_{n+1}$ are linearly independent points in an n-dimensional Euclidean space. Then the n-dimensional volume V of the n-simplex with vertices A_i satisfies the inequality*

$$V \leq \frac{\ell^n}{n! \sqrt{n^n (n+1)^{n+1}}}.$$

The equality is attained if and only if the lengths of all edges A_iA_{i+1} are equal and the simplex is regularly cyclic with the given polygon distinguished.

Proof. Consider, in the given Euclidean space, the class of all n-simplices which contain a metric net in the form of a polygon of length ℓ. By Theorem 1.1, the maximum volume among all such simplices is attained among cyclic simplices for which the polygon of the net is distinguished. The problem is thus restricted just to cyclic simplices. We show, without going to technical details, that the maximum is attained for the regular cyclic simplex.

We first use the lemma:

Lemma 2.6. *Let $A_1, \ldots, A_{n+1}$ be linearly independent points in the Euclidean n-space, let $\mathbf{a}_i = A_{i+1} - A_i$, $i = 1, \ldots, n$ be the oriented vectors of n edges of the closed polygon with these vertices. Then the square of the n-dimensional volume of the n-simplex with vertices $A_1, \ldots, A_{n+1}$ satisfies*

$$V^2 = \frac{1}{(n!)^2} \det G(\mathbf{a}_1, \ldots, \mathbf{a}_n), \tag{2}$$

where $G(\mathbf{a}_1, \ldots, \mathbf{a}_n) = (\langle \mathbf{a}_i, \mathbf{a}_j \rangle)$ is the Gram matrix of mutual inner products of the vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

To compute now the volume of the cyclic simplex from Theorem 2.3, we obtain

$$\det G(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \frac{1}{p^n} \det \begin{pmatrix} p_1(p-p_1) & -p_1p_2 & \cdots & -p_1p_n \\ -p_1p_2 & p_2(p-p_2) & \cdots & -p_2p_n \\ \cdots & \cdots & \cdots & \cdots \\ -p_1p_n & -p_2p_n & \cdots & p_n(p-p_n) \end{pmatrix},$$

where again $p = \sum_{k=1}^{n+1} p_k$.

To simplify the determinant, add the second, third, ..., n-th column to the first column in the matrix. The first column will then have entries p_1p_{n+1}, p_2p_{n+1}, ..., p_np_{n+1}; factoring out p_{n+1} from the first column and adding its p_2-multiple to the second column, p_3-multiple to the third, ..., p_n-multiple to the last column, we finally obtain

$$\det G(\mathbf{a}_1, \ldots, \mathbf{a}_n) = \frac{1}{p} p_1 p_2 \cdots p_{n+1}. \tag{3}$$

Observe now that in case B, the maximum cannot occur. Indeed, let us have a cyclic simplex Σ with the lengths ℓ_1, ℓ_2, ..., ℓ_{n+1} of the edges of the distinguished polygon, such that (1) is fulfilled.

This simplex is then also a simplex which has a Hamiltonian path with edges $\ell_1, \ell_2, \ldots, \ell_n$. By Theorem 2.1, its volume does not exceed the volume of the Schlaefli simplex with these edges; its edge joining the first and last vertex has length $\sqrt{\sum_1^n \ell_i^2}$. Since $\ell_{n+1} > \sqrt{\sum_1^n \ell_i^2}$, this Schlaefli simplex can be proportionally enlarged to have this Hamiltonian polygon of length ℓ. Therefore, the volume of the simplex Σ can be increased in the class of simplices having a Hamiltonian polygon of length ℓ.

We have thus to find the maximum of the ratio $\frac{1}{p} p_1 p_2 \cdots p_{n+1}$ under the condition that

$$\frac{1}{\sqrt{p}} [\sqrt{p_1(p-p_1)} + \sqrt{p_2(p-p_2)} + \ldots + \sqrt{p_{n+1}(p-p_{n+1})}] = \ell. \tag{4}$$

Maximizing the logarithm and using the standard technique for convex functions, one gets that the maximum is attained for the case that all the numbers p_i are equal, i. e. that all the lengths $\ell_1, \ldots, \ell_{n+1}$ are equal to $\frac{1}{n+1}\ell$.

By the formulae (2), (3), and (4), the common value p_0 of the p_i's is $p_0 = \frac{1}{n(n+1)}\ell^2$, $p = \frac{1}{n}\ell^2$, and the volume V_m of the maximal n-simplex satisfies

$$V_m^2 = \frac{\ell^{2n}}{(n!)^2 n^n (n+1)^{n+1}}.$$

The result follows. □

Let us define the *girth* of an n-simplex in a Euclidean n-space as the minimum of sums of lengths over all Hamiltonian polygons on the simplex.

Corollary 2.7. *The girth of an n-simplex with n-dimensional volume V is always greater than or equal to $\sqrt{n(n+1)}\,\sqrt[n]{n!\sqrt{n+1}V}$.*

Equality is attained if and only if the simplex is regularly cyclic.

It is immediate that the regularly cyclic 2-simplex is the equilateral triangle. The regularly cyclic 3-simplex is the tetrahedron which is obtained from a square by parallelly lifting one diagonal in the perpendicular direction to the plane so that the distance of the two new diagonals equals half of the length of each diagonal.

References

[Bor69] K. Borsuk, *Multidimensional analytic geometry*. PWN, Warsaw, 1969.

[Fie57] M. Fiedler, Über qualitative Winkeleigenschaften der Simplexe. *Czechoslovak Math. J.* 7(82) (1957), 463–478.

[Fie61] M. Fiedler, Über zyklische n-Simplexe und konjugierte Raumvielecke. *CMUC* 2,2 (1961), 3–26.

On Finitely Generated Varieties of Distributive Double p-algebras and their Subquasivarieties

Václav Koubek[1]⋆ and Jiří Sichler[2]⋆

[1] Department of Theoretical Computer Science and Mathematical Logic and Institute for Theoretical Computer Science, Faculty of Mathematics and Physics, Charles University, Malostranské nám. 25, 118 00 Praha 1, Czech Republic
koubek@kti.ms.mff.cuni.cz

[2] Department of Mathematics, University of Manitoba, Winnipeg, Manitoba, R3T 2N2 Canada
sichler@cc.umanitoba.ca

Summary. A quasivariety $\mathbb{Q}$ is Q-universal if, for any quasivariety $\mathbb{V}$ of algebraic systems of a finite similarity type, the lattice $L(\mathbb{V})$ of all subquasivarieties of $\mathbb{V}$ is isomorphic to a quotient lattice of a sublattice of the lattice $L(\mathbb{Q})$ of all subquasivarieties of $\mathbb{Q}$. We investigate Q-universality of finitely generated varieties of distributive double p-algebras. In an earlier paper, we proved that any finitely generated variety of distributive double p-algebras categorically universal modulo a group is also Q-universal. Here we consider the remaining finitely generated varieties of distributive double p-algebras and state a problem whose solution would complete the description of all Q-universal finitely generated varieties of distributive double p-algebras.

AMS Subject Classification. Primary: 06D15; Secondary: 08B15, 18B15.
Keywords. distributive double p-algebra, variety, quasivariety, categorical universality, Q-universality, critical algebra.

1 Introduction

An *algebraic system* of a finite similarity type Δ is a nonvoid set X endowed by a family of finitary operations and/or relations of the type Δ. The system is an *algebra* if the type indexes only operations, and it is a *relational system* if the type indexes only relations.

For a class $\mathbb{K}$ of algebraic systems of a given finite similarity type, let $\mathbf{Q}(\mathbb{K})$ denote the quasivariety generated by $\mathbb{K}$ (that is, the least quasivariety

⋆ The authors gratefully acknowledge the support of 1M0021620808, a project of the Czech Ministry of Education and of the NSERC of Canada.

containing $\mathbb{K}$). In general, $\mathbf{Q}(\mathbb{K}) = \mathbf{ISPP_u}(\mathbb{K})$, where the operator $\mathbf{P_u}$ closes its argument under all ultraproducts, the operator $\mathbf{P}$ produces all Cartesian products, while $\mathbf{S}$ produces all subobjects and $\mathbf{I}$ gives all isomorphic images. If $\mathbb{K}$ is a finite set of finite algebraic systems then $\mathbf{Q}(\mathbb{K}) = \mathbf{ISP}(\mathbb{K})$.

For a quasivariety $\mathbb{K}$, let $L(\mathbb{K})$ denote the inclusion-ordered lattice of all its subquasivarieties. A quasivariety $\mathbb{K}$ of algebraic systems of finite type is *Q-universal* if, for any quasivariety $\mathbb{M}$ of finite type, $L(\mathbb{M})$ is a homomorphic image of a sublattice of $L(\mathbb{K})$. Sapir introduced this notion in [Sap85], where he showed that the variety of commutative 3-nilpotent semigroups is Q-universal. Numerous quasivarieties of algebras were shown to be Q-universal (see [AD94b] and [Gor98]). For more recent results on Q-universality of small semigroup varieties, see [DK05].

For any Q-universal quasivariety $\mathbb{K}$, the free lattice on $\aleph_0$ free generators is embeddable into $L(\mathbb{K})$, and $|L(\mathbb{K})| = 2^{\aleph_0}$.

Recall that an algebra $A = (L, \vee, \wedge, ^+, ^*, 0, 1)$ of the type $(2, 2, 1, 1, 0, 0)$ is a *distributive double p-algebra* (or *dp-algebra*) if $(L, \vee, \wedge, 0, 1)$ is a distributive $(0, 1)$-lattice, and * and $^+$ are, respectively, the unary operations of pseudocomplementation and dual pseudocomplementation: the operation * is determined by the requirement that $x \leq a^*$ be equivalent to $x \wedge a = 0$, while $y \geq a^+$ is to be equivalent to $y \vee a = 1$. Together with all their homomorphisms preserving these six operations, *dp*-algebras form a variety. The *dp*-algebras satisfying $x \vee x^* \geq y \wedge y^+$ are called *regular* and they form a variety we denote $\mathbb{R}$. We note that, when combined, earlier results by W. Dziobiak [Dzi85a, Dzi85b] show that the variety $\mathbb{R}$ of regular *dp*-algebras (or of regular double Heyting algebras) is Q-universal. We discuss the Dziobiak's result (and another one of our own) in the concluding Section 7 of this paper. Neither result yields a finitely generated Q-universal subvariety of $\mathbb{R}$ (there are no such subvarieties), and hence both results are somewhat tangential.

Finitely generated Q-universal varieties of *dp*-algebras do exist, and they will be discussed in Section 3. Section 3 also presents our current results in their wider context and asks the question whose answer is necessary for a complete classification of finitely generated Q-universal varieties. We work in terms of Priestley spaces dual to *dp*-algebras from such varieties. Priestley duality is uniquely suited to structural considerations of algebras whose reducts are distributive $(0, 1)$-lattices, and we use it throughout the paper. This is why Section 2 on Priestley duality is needed to state our results.

Section 4 briefly discusses Priestley spaces of critical *dp*-algebras, while the subsequent Section 5 uses them to exhibit a type of finitely generated varieties of *dp*-algebras that are never Q-universal. Section 6 uses critical *dp*-algebras to show that the lattice $L(\mathbb{V})$ has an infinite antichain for any remaining finitely generated variety $\mathbb{V}$, and the concluding Section 7 discusses Q-universal subvarieties of $\mathbb{R}$ which are not finitely generated.

2 Priestley Duality

Here we review needed facts about Priestley duals of *dp*-algebras and of their finitely generated varieties.

Let $(X, \leq)$ be a poset. For a set $A \subseteq X$, let

$$(A] = \{x \in X \mid \exists a \in A \text{ with } x \leq a\} \text{ and } [A) = \{x \in X \mid \exists a \in A \text{ with } a \leq x\}.$$

A set $B \subseteq X$ is *decreasing* if $B = (B]$, and *increasing* if $B = [B)$. For any $A \subseteq X$, the set $(A]$ is the smallest decreasing subset of $(X, \leq)$ containing A, and $[A)$ is the smallest increasing subset of $(X, \leq)$ containing A. If (X, τ) is a topological space, we say that $A \subseteq X$ is *clopen* if it is both τ-closed and τ-open.

A triple $(X, \leq, \tau)$ is a *Priestley space* if $(X, \leq)$ is a poset and (X, τ) is a compact topological space such that for every $x \not\leq y$ there exists a clopen decreasing set $A \subseteq X$ with $y \in A$ and $x \notin A$. Let $\mathbb{P}$ denote the category of all Priestley spaces and all continuous order preserving maps between them, and let $\mathbb{D}$ be the variety of all distributive $(0,1)$-lattices and all their $(0,1)$-homomorphisms. A celebrated theorem of H. A. Priestley relates the categories $\mathbb{P}$ and $\mathbb{D}$ as follows.

Theorem 2.1 ([Pri70]). *There exist contravariant functors* $\mathbf{D} : \mathbb{P} \to \mathbb{D}$ *and* $\mathbf{P} : \mathbb{D} \to \mathbb{P}$ *such that* $\mathbf{P} \circ \mathbf{D}$ *and* $\mathbf{D} \circ \mathbf{P}$ *are naturally equivalent to the identity functors of their respective domains. Both* $\mathbf{P}$ *and* $\mathbf{D}$ *send finite objects to finite objects.* ◻

For a Priestley space X, let $\mathrm{Min}(X)$ and $\mathrm{Max}(X)$ denote the respective sets of all its minimal and maximal elements, and let $\mathrm{Mid}(X) = X \setminus (\mathrm{Min}(X) \cup \mathrm{Max}(X))$. For $Y \subseteq X$, denote $\mathrm{Min}(Y) = \mathrm{Min}(X) \cap (Y]$, $\mathrm{Max}(Y) = \mathrm{Max}(X) \cap [Y)$ and $\mathrm{Ext}(Y) = \mathrm{Min}(Y) \cup \mathrm{Max}(Y)$. When $Y = \{y\}$, we write $\mathrm{Min}(y)$, $\mathrm{Max}(y)$ and $\mathrm{Ext}(y)$ instead of $\mathrm{Min}(\{y\})$, $\mathrm{Max}(\{y\})$ and $\mathrm{Ext}(\{y\})$. If $Y \subseteq X$ is non-void, then both $\mathrm{Min}(Y)$ and $\mathrm{Max}(Y)$ are non-void.

Recall that a distributive double *p*-algebra A is *regular* if $x \vee x^* \geq y \wedge y^+$ for all $x, y \in A$.

The theorem below specifies the Priestley duality for *dp*-algebras and their homomorphisms.

Theorem 2.2 ([Pri84]). *Let* $\mathbf{P} : \mathbb{D} \to \mathbb{P}$ *be the functor from Theorem 2.1 and let* $h : L \to L'$ *be a* $\mathbb{D}$*-morphism. Then*

(1) *L is a dp-algebra if and only if* $(Y]$ *is clopen for every clopen increasing set* $Y \subseteq \mathbf{P}L$ *and* $[Y)$ *is clopen for every clopen decreasing set* $Y \subseteq \mathbf{P}L$*;*
(2) *if L and L' are dp-algebras then h is a dp-algebra homomorphism if and only if* $\mathbf{P}h(\mathrm{Min}(x)) = \mathrm{Min}(\mathbf{P}h(x))$ *and* $\mathbf{P}h(\mathrm{Max}(x)) = \mathrm{Max}(\mathbf{P}h(x))$ *for every element x of* $\mathbf{P}L'$*;*
(3) $\mathrm{Min}(\mathbf{P}L)$ *and* $\mathrm{Max}(\mathbf{P}L)$ *are closed for any dp-algebra L;*
(4) *L is a regular dp-algbera if and only if* $\mathrm{Ext}(\mathbf{P}L) = \mathbf{P}L$*;*

(5) *if h is a dp-algebra homomorphism and Z is an order component of* $\mathbf{P}L$, *then* $Z \cap \mathrm{Im}(\mathbf{P}h) \neq \emptyset$ *if and only if* $\mathrm{Ext}(Z) \subseteq \mathrm{Im}(\mathbf{P}h)$*;*
(6) *a dp-homomorphism h is injective if and only if* $\mathbf{P}h$ *is surjective;*
(7) *a dp-homomorphism h is surjective if and only if* $\mathbf{P}h$ *is a homeomorphism and order isomorphism of* $\mathbf{P}L'$ *onto a closed order subspace Z of* $\mathbf{P}L$ *satisfying* $\mathrm{Ext}(Z) \subseteq Z$*; the dp-map* $\mathbf{P}h$ *is then called a dp-*inclusion
(8) *a family* $\{h_i : L \to L_i \mid i \in I\} \subseteq \mathbb{D}$ *is separating if and only if the set* $\bigcup_{i \in I} \mathrm{Im}(\mathbf{P}h_i)$ *is dense in* $\mathbf{P}L$. ☐

A Priestley space X satisfying (1) of Theorem 2.2 is called a *dp-space.* For *dp*-spaces X_1 and X_2, a continuous order preserving mapping $f : X_1 \to X_2$ such that $\mathrm{Min}(f(x)) = f(\mathrm{Min}(x))$ and $\mathrm{Max}(f(x)) = f(\mathrm{Max}(x))$ for every $x \in X_1$ is called a *dp-map.* Thus the category $\mathbb{DP}$ of all *dp*-spaces and all *dp*-maps between them is the category dual to the variety of all *dp*-algebras. Observe also that any finite poset $(X, \leq)$ endowed with the discrete topology is a *dp*-space.

For a *dp*-algebra L, the least *dp*-subalgebra $A \subseteq L$ containing all elements x^* and x^+ with $x \in L$ and closed under the formation of relative complements (in the sense that every $x \in L$ for which $a, a \vee x, a \wedge x \in A$ must also belong to A) is called the *rudiment* of L and denoted as $R(L)$. Let r_L denote the inclusion homomorphism $R(L) \subseteq L$. A *dp*-algebra L is called *rudimentary* if $R(L) = L$. A directly indecomposable rudimentary *dp*-algebra is called a *nucleus.* We recall some basic properties of *dp*-spaces dual to these algebras.

Theorem 2.3 ([KS94, KS98b, KS04a]). *Let A be a dp-algebra and* $\mathbf{P}A = X$. *Then*

(1) *the dp-map* $\rho_X = \mathbf{P}r_A$ *is surjective, and* $(x_1, x_2) \in \mathrm{Ker}(\rho_X)$ *if and only if* $\mathrm{Min}(x_1) = \mathrm{Min}(x_2)$ *and* $\mathrm{Max}(x_1) = \mathrm{Max}(x_2)$*;*
(2) *if* $Y = \mathbf{P}(R(A))$ *and* $\rho_X = \mathbf{P}r_A$ *then* $y_1 \leq y_2$ *in Y if and only if there exist* $x_1 \in \rho_X^{-1}(y_1)$ *and* $x_2 \in \rho_X^{-1}(y_2)$ *such that* $x_1 \leq x_2$*;*
(3) *for any dp-algebra homomorphism* $f : B \to A$ *there exists a unique dp-map* $g : \mathbf{P}(R(A)) \to \mathbf{P}(R(B))$ *with* $\mathbf{P}(r_B) \circ \mathbf{P}f = g \circ \mathbf{P}(r_A)$*;*
(4) *if X is finite, order connected and satisfies* $\mathrm{Ext}(x) = \mathrm{Ext}(y)$ *only when* $x = y$*, then every dp-endomorphism of X is invertible.* ☐

Definition and notation. We say that a *dp*-space X is *rudimentary* if there exist no distinct $x, y \in X$ with $\mathrm{Ext}(x) = \mathrm{Ext}(y)$. When order connected, a finite rudimentary *dp*-space X is the *dp*-space of a nucleus, and we call X a *nucleus* as well. For any *dp*-space X, we denote

$$\mathrm{Rud}(X) = \mathbf{P}(R(\mathbf{D}(X))) \quad \text{and} \quad \rho_X = \mathbf{P}(r_{\mathbf{D}(X)}) : X \to \mathrm{Rud}(X).$$

For finite $k \geq 0$, a nucleus X is called a *k-nucleus* if k is the maximal number of non-singleton ordered pairs occuring in a single order component of $\mathrm{Mid}(X)$.

Let X be any *dp*-space. An element $x \in \mathrm{Mid}(X)$ is *min-defective* if $\rho_X(x) = \rho_X(y)$ for some $y \in \mathrm{Min}(X)$, and x is *max-defective* if $\rho_X(x) = \rho_X(y)$ for some

$y \in \mathrm{Max}(X)$. We say that x is *defective* if it is min-defective or max-defective. It is clear that a rudimentary *dp*-space cannot have any defective elements.

On any finite *dp*-space X, define an equivalence θ_X by

$$(x, y) \in \theta_X \text{ if and only if } x, y \in \mathrm{Mid}(X) \text{ and } \mathrm{Ext}(x) = \mathrm{Ext}(y) \text{ or } x = y,$$

and let $S(X) = X/\theta_X$ denote the order quotient of X modulo θ_X. Let $\sigma_X : X \to S(X)$ be the order preserving map with $\mathrm{Ker}(\sigma_X) = \theta_X$. It is then clear that $\sigma_X^{-1}\{\sigma_X(e)\} = \{e\}$ for every $e \in \mathrm{Ext}(X)$, that $\mathrm{Ker}(\sigma_X) \subseteq \mathrm{Ker}(\rho_X)$, and that $\sigma_X = \rho_X$ if and only if X has no defective elements. The *dp*-space $S(X)$ will be called the *semi-rudiment* of X and, if X is connected, then $S(X)$ is the *semi-nucleus* of X. For $k = 2, 3$, a semi-nucleus Y is a *k-semi-nucleus* if $\mathrm{Rud}(Y)$ is a 1-nucleus, Y has exactly one defective element x, and in the order component of $\mathrm{Mid}(Y)$ containing x there are exactly k distinct ordered pairs.

Finally, we recall several facts about finitely generated varieties of *dp*-algebras. First, a *dp*-algebra A with $\mathbf{P}A = X$ belongs to some finitely generated variety $\mathbb{V}$ if and only if there exists a finite k such that $|\mathrm{Ext}(C)| \leq k$ for every order component C of X, see [KS94] or [KS98b].

According to Davey [Dav78], a finite poset X is the *dp*-space of a subdirectly irreducible algebra if and only if X is order connected and $|\mathrm{Mid}(X)| \leq 1$. For any *dp*-space X dual to a *dp*-algebra from a finitely generated variety $\mathbb{V}$ and for any $x \in X$, let $C(x)$ be the order component of X containing x. Let X_x denote the (closed) ordered subspace $\mathrm{Ext}(C(x)) \cup \{x\}$ of X. Then $\mathbf{D}X_x \in \mathbb{V}$ is subdirectly irreducible for every $x \in X$.

Theorem 2.4 ([KS94, KS98b]). *For any dp-space X and for any finitely generated variety $\mathbb{V}$ of dp-algebras,*

(1) $\mathbf{D}X \in \mathbb{V}$ *if and only if* $\mathbf{D}X_x \in \mathbb{V}$ *for every* $x \in X$*;*
(2) *if* $\mathbf{D}X \in \mathbb{V}$ *then* $\mathbf{D}(\mathrm{Rud}(X)) \in \mathbb{V}$*;*
(3) $\mathbf{D}X \in \mathbb{V}$ *if and only if* $\mathbf{D}(S(X)) \in \mathbb{V}$*;*
(4) $\mathbb{V}$ *contains only finitely many pairwise non-isomorphic semi-nuclei.* ◻

We conclude this section by a simple but useful observation based on Theorem 2.4.

For a finite space X with a semi-nucleus $S(X)$, we define *extremal order* $\leq_e$ on the underlying set Y of $S(X)$ by writing $x \leq_e y$ if and only iff $x \leq y$ in $S(X)$ in case when $\{x, y\} \cap \mathrm{Ext}(S(X)) \neq \emptyset$ and, for $x, y \in \mathrm{Mid}(S(X))$ we write $x \leq_e y$ exactly when $\mathrm{Min}(x) \subseteq \mathrm{Min}(y)$ and $\mathrm{Max}(x) \supseteq \mathrm{Max}(y)$. The *discrete order* $\leq_d$ on Y is obtained by setting $x \leq_d y$ iff $x \leq y$ in $S(X)$ for $\{x, y\} \cap \mathrm{Ext}(S(X)) \neq \emptyset$ and $x \leq_d y$ iff $x = y$ for $x, y \in \mathrm{Mid}(S(X))$. Then $(Y, \leq_e)$ and $(Y, \leq_d)$ are semi-nuclei and the identity map 1_Y on the set Y is the Priestley dual of injective *dp*-homomorphisms $\mathbf{D}(Y, \leq_e) \to \mathbf{D}S(X)$ and $\mathbf{D}S(X) \to \mathbf{D}(Y, \leq_d)$, see Theorem 2.2(6). By Theorem 2.4, any finitely generated variety $\mathbb{V}$ either contains $\mathbf{D}X$ along with all semi-nuclei $\mathbf{D}S$ such that Y is the underlying set of S and the the order of S contains $\leq_d$ and is contained in $\leq_e$ or $\mathbb{V}$ contains none of these algebras.

3 Summary of Earlier and Present Results

This paper is a step towards a full characterization of all finitely generated Q-universal varieties of *dp*-algebras.

A concrete category $\mathbb{K}$ is (*algebraically*) *universal* if there exists a full and faithful functor $F : \mathbb{GRA} \to \mathbb{K}$ from the category $\mathbb{GRA}$ of all graphs and all their compatible mappings into $\mathbb{K}$, and $\mathbb{K}$ is *finite-to-finite universal* if such a functor sends finite graphs to finite $\mathbb{K}$-objects.

In [AD01], M. E. Adams and W. Dziobiak proved that any finite-to-finite universal quasivariety $\mathbb{K}$ of algebraic systems is Q-universal. All (finite-to-finite) universal finitely generated varieties of *dp*-algebras were structurally characterized in [KS94] (in [Kou90]) as the varieties containing a 2-nucleus (a 3-nucleus) X whose only endomorphism extending the identity map $1_{\mathrm{Mid}(X)}$ of $\mathrm{Mid}(X)$ is the identity endomorphism 1_X of X itself. The combination of the results from [AD01] and [Kou90] thus gives numerous examples of finitely generated Q-universal varieties of *dp*-algebras.

Our recent paper [KS06$^+$] proves the following extension of the latter result.

Theorem 3.1 ([KS06$^+$]). *Any finitely generated variety $\mathbb{V}$ of dp-algebras containing a 2-nucleus is Q-universal.* □

On the other hand, [KS04b] shows many failures of Q-universality.

Theorem 3.2 ([KS04b]). *For any finitely generated variety $\mathbb{V}$ of dp-algebras, the lattice $L(\mathbb{V})$ of all its subquasivarieties is finite if and only if $\mathbb{V}$ contains no 1-nucleus.* □

In view of these two theorems, it is natural to ask about the subquasivariety lattices $L(\mathbb{V})$ of finitely generated varieties $\mathbb{V}$ containing a 1-nucleus but no 2-nucleus. In particular, one may ask the following question.

Problem 3.3. Must every Q-universal finitely generated variety $\mathbb{V}$ of *dp*-algebras contain a 2-nucleus?

The present paper adresses this problem by investigating varieties containing a 1-nucleus but no 2-nucleus. Perhaps surprisingly, properties of such varieties appear to depend on the semi-nuclei they contain. We prove the two results below.

Theorem 3.4. *If a finitely generated variety $\mathbb{V}$ of dp-algebras contains no 2-nucleus and no 3-semi-nucleus, then $L(\mathbb{V})$ is countable (and hence $\mathbb{V}$ is not Q-universal).*

Theorem 3.5. *If a finitely generated variety $\mathbb{V}$ of dp-algebras contains a 3-semi-nucleus, then $L(\mathbb{V})$ has a countably infinite antichain.*

Even though Theorem 3.5 appears to speak (somewhat inconclusively) for the opposite, we still conjecture that the answer to Problem 3.3 is positive.

4 On Critical Algebras

We recall that a finite algebra A is *critical* if it is not a subdirect product of its proper subalgebras. From [KS04b] we now recall a non-structural characterization of the duals of critical *dp*-algebras. We call such spaces *critical.*

Proposition 4.1 ([KS04b]). *A finite dp-space X is critical if and only if there is some $x \in X$ such that $x \notin \mathrm{Im}(f)$ for every non-invertible dp-endomorphism f of X.* □

For a finite *dp*-space X, denote

$$\mathcal{F} = \mathrm{End}(X) \setminus \mathrm{Aut}(X) \quad \text{and} \quad \mathcal{F}(X) = \bigcup\{\mathrm{Im}(f) \mid f \in \mathcal{F}\}.$$

Thus X is critical exactly when $\mathcal{F}(X) \neq X$.

Next, we say that a preorder P is *almost discrete* if there exists at most one $x_0 \in P$ such that $x_0 \geq x$ for some $x \in P \setminus \{x_0\}$. When such P is not an antichain, then x_0 is uniquely determined, and we say that x_0 is *exceptional* in P. It is clear that any almost discrete preorder is, in fact, a partial order.

Let $\mathrm{Comp}(X)$ denote the set of all order components of a finite *dp*-space X. For $C_1, C_2 \in \mathrm{Comp}(X)$, we write $C_1 \leq C_2$ if there exists a *dp*-map $f : C_2 \to C_1$. Then $\leq$ is a preorder on $\mathrm{Comp}(X)$, and we thus speak about exceptional components. If $C_1, C_2 \in \mathrm{Comp}(X)$ with $C_1 \leq C_2 \neq C_1$, then there exists a *dp*-map $f : C_2 \to C_1$. Define $g_f : X \to X$ by

$$g_f(x) = \begin{cases} x & \text{if } x \in X \setminus C_2, \\ f(x) & \text{if } x \in C_2. \end{cases}$$

Then $g_f \in \mathrm{End}(X)$ is idempotent and $\mathrm{Im}(g_f) = X \setminus C_2$. Thus $g_f \in \mathcal{F}$. Next, if $x \in X$ is a min-defective element covering exactly one $y \in X$, we define $h_x : X \to X$ by

$$h_x(z) = \begin{cases} z & \text{if } z \neq x, \\ y & \text{if } z = x. \end{cases}$$

It is clear that $h_x \in \mathrm{End}(X)$ is idempotent with $\mathrm{Im}(f) = X \setminus \{x\}$. If $x \in X$ is max-defective such that x is covered by exactly one element then h_x is analogously defined. Using these *dp*-maps, we easily obtain the following lemma which extends certain claims made in [KS04b].

Lemma 4.2. *Let X be a finite critical dp-space. Then*

(1) *the preorder of* $\mathrm{Comp}(X)$ *is almost discrete;*
(2) *X has at most one defective element;*
(3) *if $x \in X$ is defective and $C \in \mathrm{Comp}(X)$ is exceptional then $x \in C$.*

Proof. For (1), let $C_1, C_2, C_3, C_4 \in \mathrm{Comp}(X)$ be such that $C_1 \neq C_3$, $C_2 < C_1$ and $C_4 < C_3$. Then there exist *dp*-maps $f_1 : C_1 \to C_2$ and $f_2 : C_3 \to C_4$ and hence g_{f_1} and g_{f_2} are non-invertible *dp*-endomorphisms of X with $\mathrm{Im}(g_{f_1}) \cup$

$\mathrm{Im}(g_{f_2}) = X$. This contradicts Proposition 4.1, and hence $\mathrm{Comp}(X)$ is almost discrete. To prove (2), let X have at least two defective elements. Then, by the finiteness of X, there exist distinct defective elements $x, y \in X$ such that if $z \in \{x, y\}$ is min-defective then z covers exactly one element, if $z \in \{x, y\}$ is max-defective then it is covered exactly by one element. Then h_x and h_y are non-invertible *dp*-endomorphisms of X with $\mathrm{Im}(h_x) \cup \mathrm{Im}(h_y) = X$. This again contradicts Proposition 4.1, so that X has at most one defective element. Finally, if $C_1 < C_2$ in $\mathrm{Comp}(X)$ and $f : C_2 \to C_1$ is a *dp*-map, and if x is the defective element of X, then $g_f, h_x \in \mathcal{F}$ and $\mathrm{Im}(g_f) \cup \mathrm{Im}(h_x) = X \setminus (C_2 \cap \{x\}) \subseteq \mathcal{F}(X)$. By Proposition 4.1, we must have $\mathcal{F}(X) \neq X$, and this is possible only when $x \in C_2$. This proves (3). ◻

5 Varieties Containing no 2-nuclei and no 3-semi-nuclei

Here we prove Theorem 3.4, by means of describing all critical algebras from finitely generated varieties with no 2-nuclei and no 3-semi-nuclei. In this section, we call such varieties *nearly regular*.

Let $(P, \leq)$ be a finite poset. A subset $Z = \{z_0, z_1, \ldots, z_n\}$ of $(P, \leq)$ is a *zigzag* if either $z_{2i} > z_{2i+1} < z_{2i+2}$ for all $i = 0, 1, \ldots, \lfloor \frac{n-2}{2} \rfloor$ and if n is odd then $z_{n-1} > z_n$ or $z_{2i} < z_{2i+1} > z_{2i+2}$ for all $i = 0, 1, \ldots, \lfloor \frac{n-2}{2} \rfloor$ and if n is odd then $z_{n-1} < z_n$, and P has no other non-singleton ordered pairs within Z. For $u, v \in P$, let $\mathrm{dist}(u, v)$ be the least number n such that there exists a sequence $u = u_0, u_1, \ldots, u_n = v$ such that u_i and u_{i+1} are comparable for all $i = 0, 1, \ldots, n-1$. Then $u_0, u_1, \ldots, u_n$ is called a *shortest sequence* connecting u and v. Observe that any such sequence must be a zigzag. For nonvoid $U, V \subseteq P$, define $\mathrm{dist}(U, V) = \min\{\mathrm{dist}(u, v) \mid u \in U, v \in V\}$, and write $\mathrm{dist}(U, V) = \infty$ if no sequence connects any $u \in U$ to any $v \in V$.

We begin with two simple claims, see also [Kou85].

Lemma 5.1. *Let $(P, \leq)$ be a poset. If U and V are non-void, disjoint subsets of P such that U is either increasing or decreasing and* $\mathrm{dist}(U, V) = n < \infty$ *then for $u \in U$ and $v \in V$ with* $\mathrm{dist}(U, V) = \mathrm{dist}(u, v)$ *and for a shortest sequence $u = u_0, u_1, \ldots, u_n = v$ there exists an order preserving idempotent mapping $f : P \to P$ such that* $\mathrm{Im}(f) = \{u_i \mid 0 \leq i \leq n\}$, $f(U) = \{u\}$ *and one of these four cases occurs:*

(1) *if $[U) = U$ and if n is even then $u = u_0 > u_1$ and $f([V)) = \{v\}$;*
(2) *if $[U) = U$ and if n is odd then $u = u_0 > u_1$ and $f((V]) = \{v\}$;*
(3) *if $(U] = U$ and if n is odd then $u = u_0 < u_1$ and $f([V)) = \{v\}$;*
(4) *if $(U] = U$ and if n is even then $u = u_0 < u_1$ and $f((V]) = \{v\}$.*

Proof. Let $u = u_0, u_1, \ldots, u_n = v$ be a shortest sequence between U and V. Assume that $[U) = U$ and n is even. If $u_1 > u_0$ then $u_1 \in U$ and $\mathrm{dist}(U, V) \leq \mathrm{dist}(u_1, v) < \mathrm{dist}(u, v) = \mathrm{dist}(U, V)$ and this contradicts the hypothesis.

Thus $u_1 < u_0$ and hence $u_{2i} > u_{2i+1} < u_{2i+2}$ for all $i = 0, 1, \ldots, n/2 - 1$. Let us define $V_0 = [V)$ and, once V_{2i} is defined, we set $V_{2i+1} = (V_{2i}] \setminus \bigcup_{j=0}^{2i} V_j$ and $V_{2i+2} = [V_{2i+1}) \setminus \bigcup_{j=0}^{2i+1} V_j$. Then $V_j \cap V_k = \emptyset$ whenever $j \neq k$, and $u_{n-j} \in V_j$ for all $j = 0, \ldots, n$. Furthermore, the sets V_{2i} are increasing and all V_{2i+1} are decreasing. From their definition we obtain $\bigcup_{j=0}^{2i+1} V_j = (V_{2i}] \cup \bigcup_{j=0}^{2i+1} V_j$, that is, $(V_{2i}] \subseteq \bigcup_{j=0}^{2i+1} V_j$ and also $[V_{2i+1}) \subseteq \bigcup_{j=0}^{2i+2} V_j$. A simple induction using these facts and beginning with the inclusions $(V_0] \subseteq V_0 \cup V_1$ and $[V_1) \subseteq V_0 \cup V_1 \cup V_2$ shows that

$$[V_{2i+1}) \subseteq V_{2i} \cup V_{2i+1} \cup V_{2i+2} \text{ and } (V_{2i}] \subseteq V_{2i-1} \cup V_{2i} \cup V_{2i+1} \tag{v}$$

for all $i = 0, 1, \ldots$ (where we set $V_{-1} = \emptyset$). It follows that each $\bigcup_{j=0}^{2i} V_j$ is increasing and each $\bigcup_{j=0}^{2i+1} V_j$ is decreasing.

Since $\operatorname{dist}(U, V) = n$ is even, we have $\operatorname{dist}(U, [V)) = n$, and hence $U \cap V_j = \emptyset$ for all $j = 0, 1, \ldots, n-1$. For $x \in P$ define

$$f(x) = \begin{cases} u_i & \text{if } x \in V_{n-i} \text{ for } i = 0, 1, \ldots, n-1, \\ u_0 & \text{if } x \in P \setminus \bigcup_{i=0}^{n-1} V_i. \end{cases}$$

Thus $f : P \to P$ is a well-defined map whose image is the zigzag $u = u_0 > u_1 < \ldots > u_{n-1} < u_n = v$. Since n is even, the set $P \setminus \bigcup_{i=0}^{n-1} V_i$ is increasing, and from (v) it follows that f is an order preserving idempotent map with $f(U) = \{u\}$ and $f([V)) = \{v\}$. This completes the proof of (1). The remaining three claims are proved similarly. □

Lemma 5.2. *Let $(P, \leq)$ be a poset and let U and V be nonvoid subsets of P with $\operatorname{dist}(U, V) = \infty$. Then for every $u \in U$ and $v \in V$ there exists an order preserving idempotent mapping $f : P \to P$ with $\operatorname{Im}(f) = \{u, v\}$, $f(U) = \{u\}$ and $f(V) = \{v\}$.*

Proof. Since $\operatorname{dist}(U, V) = \infty$, the union W of all order components intersecting U is both increasing and decreasing and does not intersect V. For any given $u \in U$ and $v \in V$ set

$$f(x) = \begin{cases} v & \text{if } x \in P \setminus W, \\ u & \text{if } x \in W. \end{cases}$$

Then $f : P \to P$ is an order preserving idempotent mapping with $\operatorname{Im}(f) = \{u, v\}$, $f(U) = \{u\}$ and $f(V) = \{v\}$. □

In view of Theorem 2.4 and the remark following it, the Priestley dual X of any algebra from a nearly regular variety has these two properties:

(nr1) any order component of $\operatorname{Mid}(\operatorname{Rud}(X))$ has at most two elements;

(nr2) any defective $x \in \mathrm{Mid}(X)$ and non-defective $y, z \in \mathrm{Mid}(X)$ satisfy either $\mathrm{Min}(x) \subseteq \mathrm{Min}(y) \subseteq \mathrm{Min}(z)$ and $\mathrm{Max}(z) \subseteq \mathrm{Max}(y) \subseteq \mathrm{Max}(x)$, or $\mathrm{Min}(z) \subseteq \mathrm{Min}(y) \subseteq \mathrm{Min}(x)$ and $\mathrm{Max}(x) \subseteq \mathrm{Max}(y) \subseteq \mathrm{Max}(z)$ only when $\mathrm{Ext}(y) = \mathrm{Ext}(z)$.

Any *dp*-space X satisfying (nr1) and (nr2) is called *nearly regular.*

Given a *dp*-space X and the *dp*-map $\rho_X : X \to \mathrm{Rud}(X)$ onto its rudiment $\mathrm{Rud}(X)$, for any $x \in X$ denote

$$E(x) = \rho_X^{-1}\{\rho_X(x)\} = \{y \in X \mid \mathrm{Ext}(y) = \mathrm{Ext}(x)\}.$$

To describe nearly regular critical *dp*-spaces, we need the following notion.

Definition 5.3. *A nearly regular dp-space X having a unique defective element x is an x-*anchored pseudorudiment *if for any non-defective $y \in \mathrm{Mid}(X)$ with $|E(y)| > 1$,*

- *the set $E(y)$ is either a a zigzag $\{u = y_0, y_1, \ldots, y_n = v\}$ or a two-element antichain $\{u, v\}$;*
- *x is comparable to $u \in E(y)$ but not to any other element of $E(y)$, and any element comparable to u belongs to the set $\mathrm{Ext}(y) \cup E(y) \cup \{x\}$;*
- *there exists a non-defective element $z \in \mathrm{Mid}(X)$ with $|E(z)| = 1$ comparable to $v \in E(y)$ but not to any other element of $E(y)$, and any element comparable to v belongs to the set $\mathrm{Ext}(y) \cup E(y) \cup \{z\}$;*
- *if $E(y) = \{u = y_0, y_1, \ldots, y_n = v\}$ is a zigzag then, for every $i = 1, 2, \ldots, n-1$, any element comparable to y_i belongs to $\mathrm{Ext}(y) \cup E(y)$.*

Theorem 5.4. *Let X be a finite nearly regular dp-space, and let $x \in X$ be defective. Then X is critical if and only if*

(1) *the preorder of $\mathrm{Comp}(X)$ is almost discrete;*
(2) *x is the only defective element of X;*
(3) *if $C \in \mathrm{Comp}(X)$ is exceptional then $x \in C$;*
(4) *X is an x-anchored pseudorudiment.*

Proof. Let X be critical and let $x \in X$ be defective. Then x is the only defective element of X and (1)–(3) hold because of Lemma 4.2. If x is min-defective then x covers a unique element of $\mathrm{Min}(X)$, if x is max-defective then x is covered by a unique element of $\mathrm{Max}(X)$, and therefore h_x is a non-invertible *dp*-endomorphism of X with $\mathrm{Im}(h_x) = X \setminus \{x\}$.

Suppose that $y \in X$ is non-defective and $|E(y)| > 1$.

Case 1. Assume that a non-defective $z \in \mathrm{Mid}(X)$ is comparable to an element of $E(y)$ only when $z \in E(y)$. Select $y' \in E(y)$ so that x is comparable to y' whenever $E(y)$ has such an element y', and make an arbitrary selection of $y' \in E(y)$ otherwise. Then the mapping $g : X \to X$ given by

$$g(u) = \begin{cases} u & \text{if } u \in X \setminus E(y), \\ y' & \text{if } u \in E(y) \end{cases}$$

is a non-invertible member of $\mathcal{F}$ such that $x \in \mathrm{Im}(g)$. But then $\mathcal{F}(X) = X$, and this contradicts Proposition 4.1. We conclude that $E(y)$ is a singleton whenever the order component of $\mathrm{Mid}(\mathrm{Rud}(X))$ containing $\rho_X(y)$ is a singleton.

Case 2. Assume that a non-defective $z \in \mathrm{Mid}(X)$ with $E(z) \cap E(y) = \emptyset$ is comparable to an element of $E(y)$. Since X is nearly regular, the defective element x can be comparable to an element of $E(y)$ or to an element of $E(z)$ but never to both.

Case 2.1. Let there be a comparable pair $\{y', z'\}$ with $y' \in E(y)$ and $z' \in E(z)$ such that $\{x, y'\}$ is a comparable pair, or else let x be incomparable to all elements of $E(y) \cup E(z)$. In either case, the mapping $g : X \to X$ given by

$$g(u) = \begin{cases} u & \text{if } u \in X \setminus (E(y) \cup E(z)), \\ y' & \text{if } u \in E(y), \\ z' & \text{if } u \in E(z) \end{cases}$$

is an idempotent *dp*-endomorphism of X for which $x \in \mathrm{Im}(g)$. But then g must be invertible because of Proposition 4.1, contradicting the fact that $|E(y)| > 1$. This leaves us with the final

Case 2.2. The set U of all elements of $E(y)$ comparable to x and the set V of all elements of $E(y)$ comparable to an element of $E(z)$ are nonvoid and disjoint. Since both are monotone in $E(y)$, Lemma 5.1 or 5.2 respectively imply the existence of an order preserving mapping $g' : E(y) \to E(y)$ such that

(a) if $\mathrm{dist}(U, V)$ is finite, then $\mathrm{Im}(g')$ is a shortest zigzag sequence from U to V, connecting some $u \in U$ and $v \in V$;
(b) if $\mathrm{dist}(U, V) = \infty$, then $\mathrm{Im}(g') = \{u, v\}$ with $u \in U$ and $v \in V$.

Denote $\{u\} = \mathrm{Im}(g') \cap U$ and $\{v\} = \mathrm{Im}(g') \cap V$, and select some $z' \in E(z)$ comparable to v. Define $g : X \to X$ by

$$g(t) = \begin{cases} t & \text{if } t \in X \setminus (E(y) \cup (E(z)), \\ z' & \text{if } t \in E(z), \\ g'(t) & \text{if } t \in E(y). \end{cases}$$

Then g is an idempotent *dp*-endomorphism of X and $x \in \mathrm{Im}(g)$. Since $\mathcal{F}(X) \neq X$ by the hypothesis, g must be invertible, and hence the identity map. Thus $E(z) = \{z'\}$ and $E(y) \subseteq \mathrm{Im}(g)$, and $u, v \in E(y)$ are the only elements of $E(y)$ comparable to elements of $\mathrm{Mid}(X) \setminus E(y)$, namely to the respective singletons $\{x\}$ and $\{z'\}$. This proves (4).

For the converse, let a finite *dp*-space X satisfy (1)–(4). Suppose that $x \in \mathrm{Im}(f)$ for some *dp*-endomorphism of X. From (1) and (3) it follows that $f(C) \subseteq C$ for every component C of X, so that the restriction of f to $\mathrm{Ext}(X)$ is an automorphism of $\mathrm{Ext}(X)$. Several consequences follow from this fact. First we claim that $f(x) = x$. To see this, suppose that x is min-defective, let $m \in \mathrm{Min}(X)$ be such that $\mathrm{Ext}(m) = \mathrm{Ext}(x)$, and let $f(t) = x$. Then $t \in \mathrm{Mid}(X)$, and since f is injective on $\mathrm{Min}(X)$, the set $\mathrm{Min}(t) = \{n\}$ is a singleton. Thus $f(n) = m$ and $f(\mathrm{Max}(n)) = \mathrm{Max}(m) = \mathrm{Max}(x)$. Also $f(\mathrm{Max}(t)) = \mathrm{Max}(x)$, and $\mathrm{Ext}(t) = \mathrm{Ext}(n)$ follows because f is injective on $\mathrm{Max}(X)$. Therefore t is min-defective, and from (2) it follows that $t = x$. If x is max-defective the proof is analogous.

Next, since X is finite, for some $k \geq 1$ the restriction of $h = f^k$ to $\mathrm{Ext}(X)$ is the identity map of $\mathrm{Ext}(X)$, so that $h(E(t)) \subseteq E(t)$ also for every non-defective $t \in \mathrm{Mid}(X)$. Thus $h(t) = t$ whenever $E(t)$ is a singleton. If $E(y)$ is not a singleton then, by (4), there are two uniquely determined elements u and v of $E(y)$ respectively comparable to x and to a non-defective $z \notin E(y)$ with $E(z) = \{z\}$. But $h(x) = x$, $h(z) = z$ and $h(E(y)) \subseteq E(y)$, and from (4) it follows that h is the identity map on any $E(y)$ which is not a singleton. Thus f is invertible, and hence $x \notin \mathrm{Im}(f')$ for every $f' \in \mathcal{F}$. Therefore X is critical, by Proposition 4.1, and the proof is complete. □

Now we turn to nearly regular critical *dp*-spaces without defective elements.

Theorem 5.5. *A finite nearly regular dp-space X with no defective element is critical if and only if*

(1) $\mathrm{Comp}(X)$ *is almost discrete;*
(2) *at most one set $E(x)$ is not a singleton;*
(3) *if $E(x)$ is not a singleton and if $C \in \mathrm{Comp}(X)$ is exceptional, then $E(x) \subseteq C$;*
(4) *the set $E(x)$ which is not a singleton is either a zigzag $\{x_0, x_1, \ldots, x_k = z\}$ or an antichain $\{x_0, z\}$ such that each $u \in E(x) \setminus \{x_0\}$ is comparable only to the elements of $\mathrm{Ext}(x) \cup E(x)$, while there exists a unique $y \in \mathrm{Mid}(X) \setminus E(x)$ comparable to x_0 and this element satisfies $E(y) = \{y\}$.*

Proof. Let X be a nearly regular critical *dp*-space without a defective element. By Lemma 4.2, $\mathrm{Comp}(X)$ is almost discrete, so that (1) holds.

First we define some special *dp*-endomorphisms of X. If $x \in \mathrm{Mid}(X)$ and $\{\rho_X(x)\}$ is an order component of $\mathrm{Mid}(\mathrm{Rud}(X))$, we define

$$f_x(w) = \begin{cases} w & \text{if } w \in X \setminus E(x), \\ x & \text{if } w \in E(x). \end{cases}$$

If $\{\rho_X(x), \rho_X(y)\}$ is a two-element order component of $\mathrm{Mid}(\mathrm{Rud}(X))$ and if further $x \in \mathrm{Mid}(X)$, we set

$$U_x = \{u \in E(x) \mid \exists v \in E(y) \text{ comparable to } u\}.$$

By Lemmas 5.1 and 5.2, there exists an order preserving mapping $g_x : E(x) \to E(x)$ such that $u_x \in U_x$ and

$$\text{Im}(g_x) = \begin{cases} \{x\} & \text{if } x \in U_x \text{ (then } u_x = x), \\ \{x, u_x\} & \text{if } \text{dist}(\{x\}, U_x) = \infty, \\ \{x = x_0, x_1, \ldots, x_k = u_x\} & \text{if } \text{dist}(\{x\}, U_x) = k < \infty \end{cases}$$

(in the third case above, $x_0, x_1, \ldots, x_k$ is a zigzag). Observe that $\text{Im}(g_x) \cap U_x = \{u_x\}$. Define

$$f_x(w) = \begin{cases} w & \text{if } w \in X \setminus (E(x) \cup E(y)), \\ g_x(w) & \text{if } w \in E(x), \\ v_x & \text{if } w \in E(y) \end{cases}$$

where $v_x \in E(y)$ is comparable to u_x. In both cases, f_x is a dp-endomorphism of X. Clearly, f_x is invertible if and only if either $\{\rho_X(x)\}$ is an order component of $\text{Mid}(\text{Rud}(X))$ and $|E(x)| = 1$ or $\{\rho_X(x), \rho_X(y)\}$ is a two-element order component of $\text{Mid}(\text{Rud}(X))$, $|E(y)| = 1$ and $E(x) = \text{Im}(g_x)$.

We continue with three simple observations.

(a) if $x \in \text{Mid}(X)$ is such that $\{\rho_X(x)\}$ is an order component of $\text{Mid}(\text{Rud}(X))$, then $E(x)$ is a singleton;
(b) there is at most one component $\{\rho_X(x), \rho_X(y)\}$ of $\text{Mid}(\text{Rud}(X))$ such that $E(x) \cup E(y)$ has more than two elements;
(c) if $\{\rho_X(x), \rho_X(y)\}$ is a two-element order component of $\text{Mid}(\text{Rud}(X))$ and $|E(x)| > 1$, then $|E(y)| = 1$, and either $E(x) = \{x_0, x_1, \ldots, x_k\}$ is a zigzag or $E(x) = \{x_0, z\}$ is a two-element antichain and x_0 is the unique element of $E(x)$ comparable to y.

To see (a), observe that $X = \bigcup\{\text{Im}(f_{x'}) \mid x' \in E(x)\}$, and Proposition 4.1 completes the proof because X is critical. If $\{\rho_X(x), \rho_X(y)\}$ and $\{\rho_X(x'), \rho_X(y')\}$ are disjoint two-element order components of $\text{Mid}(\text{Rud}(X))$, then $X = \text{Im}(f_x) \cup \text{Im}(f_{x'})$ and Proposition 4.1 completes the proof of (b). If $\{\rho_X(x), \rho_X(y)\}$ is a two-element order component of $\text{Mid}(\text{Rud}(X))$ then $X = \bigcup_{w \in E(x) \cup E(y)} \text{Im}\, f_w$ and, by Proposition 4.1, there exists some $w \in E(x) \cup E(y)$ such that f_w is invertible because X is critical. Hence (c) follows, and (2) and (4) hold.

To prove (3) assume that $h' : C_1 \to C_2$ is a dp-map between distinct order components of X and $x \in X \setminus C_1$ with $|E(x)| > 1$. With no loss of generality, we can assume that $|\text{Im}(f_x) \cap E(x)| = 1$ (thus f_x is non-invertible). Then $X = \text{Im}(h) \cup \text{Im}(f_x)$ for

$$h(w) = \begin{cases} w & \text{if } w \in X \setminus C_1, \\ h'(w) & \text{if } w \in C_1, \end{cases}$$

and this contradiction, by Proposition 4.1, completes the proof of (3).

For the converse, suppose that X satisfies (1)–(4). First assume that $|E(x)| = 1$ for all $x \in \text{Mid}(X)$. If the preorder of $\text{Comp}(X)$ is discrete then any dp-endomorphism f of X preserves every order component and hence f is invertible because $|E(x)| = 1$ for all $x \in \text{Mid}(X)$. Thus $\mathcal{F} = \emptyset$ and

X is critical, by Proposition 4.1. If $C \in \mathrm{Comp}(X)$ is exceptional then any dp-endomorphism f either preserves all components (and f is then invertible) or else $\mathrm{Im}(f) \subseteq X \setminus C$ and, by Proposition 4.1, X is critical.

Assume that there exists $x \in \mathrm{Mid}(X)$ with $|E(x)| > 1$. By (2), the set $E(x)$ is determined uniquely and, by (4), either $E(x) = \{x_0, x_1, \ldots, x_k = z\}$ is a zigzag or $E(x) = \{x_0, z\}$ is a two-element antichain, and there exists $y \in X \setminus E(x)$ such that x_0 is the unique element of $E(x)$ comparable to y. Suppose that f is a dp-endomorphism of X with $f(w) = z$ for some $w \in X$. By (1) and (3), $f(C) \subseteq C$ for all $C \in \mathrm{Comp}(X)$ and hence f is bijective on $\mathrm{Ext}(X)$. By Theorem 2.3(3), there exists a dp-endomorphism g of $\mathrm{Rud}(X)$ with $\rho_X \circ f = g \circ \rho_X$. Since f is invertible on $\mathrm{Ext}(X)$ we conclude that g is invertible on $\mathrm{Ext}(\mathrm{Rud}(X))$ and thus g is invertible. From $g(\rho_X(w)) = \rho_X(x)$ we then obtain the existence of some $w' \in E(w)$ and some $t' \in X \setminus E(w)$ comparable to w' such that $g(\rho_X(t')) = \rho_X(y)$. Since $|E(y)| = 1$ we deduce that $f(t') = y$ and $f(w') = x_0$ and thus $|E(w)| > 1$. By (2), $E(w) = E(x)$, and hence $f(E(x)) \subseteq E(x)$ and $f(y) = y$. From (4) it follows that $f(x_0) = x_0$ and, by an easy induction, $f(u) = u$ for all $u \in E(x)$ because $z \in f(E(x))$. Hence f is invertible. Therefore $\mathcal{F}(X) = X \setminus \{z\}$, and Proposition 4.1 completes the proof. □

This completes the description of all nearly regular critical dp-spaces.

Recall that the dp-map $\sigma_X : X \to S(X)$ of X onto its semi-rudiment $S(X)$ satisfies $\sigma_X^{-1}\{\sigma_X(e)\} = \{e\}$ for every $e \in \mathrm{Ext}(X)$ and that, for any $x, y \in \mathrm{Mid}(X)$, we have $\sigma_X(x) \leq \sigma_X(y)$ if and only if $x' \leq y'$ for some $x' \in \sigma_X^{-1}\{\sigma_X(x)\}$ and $y' \in \sigma_X^{-1}\{\sigma_X(y)\}$.

Proposition 5.6. *Let X_1 and X_2 be critical nearly regular spaces such that*

(1) *$S(X_1) = S(X_2) = S$;*
(2) *for every $s \in S$, the poset $\sigma_{X_1}^{-1}\{s\}$ is order connected if and only if $\sigma_{X_2}^{-1}\{s\}$ is order connected;*
(3) *$|\sigma_{X_1}^{-1}\{s\}| \geq |\sigma_{X_2}^{-1}\{s\}|$ for every $s \in S$.*

Then there exists a surjective dp-map $f : X_1 \to X_2$.

Proof. Since both σ_{X_i} are bijective on $\mathrm{Ext}(X_i)$ and separate $\mathrm{Ext}(X_i)$ from $\mathrm{Mid}(X_i)$, there is an invertible dp-map $g : \mathrm{Ext}(X_1) \to \mathrm{Ext}(X_2)$ such that $\sigma_{X_2}(g(e)) = \sigma_{X_1}(e)$ for every $e \in \mathrm{Ext}(X_1)$. Now for any $s \in \mathrm{Mid}(S)$ and $x_i \in \sigma_{X_i}^{-1}\{s\}$ with $i = 1, 2$, we have $(\sigma_{X_2} \circ g)(\mathrm{Ext}(x_1)) = \mathrm{Ext}(s) = \sigma_{X_2}(\mathrm{Ext}(x_2))$, and hence $g(\mathrm{Ext}(x_1)) = \mathrm{Ext}(x_2)$ because σ_{X_2} is injective on $\mathrm{Ext}(X_2)$. Thus any extension $f : X_1 \to X_2$ of g to all of X_1 satisfying $\sigma_{X_2} \circ f = \sigma_{X_1}$ will have the dp-property. Hence it remains to construct a surjective order preserving map $h : \mathrm{Mid}(X_1) \to \mathrm{Mid}(X_2)$ satisfying $\sigma_{X_2} \circ h = \sigma_{X_1}$.

For any $s \in \mathrm{Mid}(S)$ such that $\sigma_{X_2}^{-1}\{s\} = \{x_2\}$ is a singleton and for any $x \in \sigma_{X_1}^{-1}\{s\}$ we set $h(x) = x_2$. In particular, if S has a defective element s, then each X_i has a unique defective element d_i and $h(d_1) = d_2$. With no loss of generality, we may assume that both d_1 and d_2 are min-defective.

Suppose that X_1 and X_2 have such defective points and that $\sigma_{X_2}^{-1}\{s\}$ is not a singleton. Then $\sigma_{X_1}^{-1}\{s\}$ is not a singleton, by (3). By Theorem 5.4, for both $i = 1, 2$, exactly one element $u_i \in \sigma_{X_i}^{-1}\{s\}$ is comparable to d_i and exactly one element $v_i \in \sigma_{X_i}^{-1}\{s\}$ is comparable to a unique element $z_i \in \mathrm{Mid}(X_i)$ with $\sigma_{X_i}(z_i) \neq s$, and $\sigma_{X_i}(z_i)$ is incomparable to $\sigma_{X_1}(d_1) = \sigma_{X_2}(d_2)$. We set $h(u_1) = u_2$ and $h(v_1) = v_2$. Since we already have $h(d_1) = d_2$ and $h(z_1) = z_2$, it follows that h sends each ordered pair with exactly one element in $\sigma_{X_1}^{-1}\{s\}$ to a similar ordered pair in X_2. By (2) and Theorem 5.4, for both $i = 1, 2$ the sets $\sigma_{X_i}^{-1}\{s\}$ are either zigzags from u_i to v_i or doubletons $\{u_i, v_i\}$, and $d_i < u_i$ and $z_i < v_i$ because d_i are min-defective and the spaces X_i are nearly regular. Since any zigzag whose endpoints u, v are maximal is an order quotient of a longer such zigzag via a map that preserves the endpoints, in either case there is a surjective order preserving map $h_s : \sigma_{X_1}^{-1}\{s\} \to \sigma_{X_2}^{-1}\{s\}$ with $h_s(u_1) = u_2$ and $h_s(v_1) = v_2$. We then define $h(x) = h_s(x)$ for every $x \in \sigma_{X_1}^{-1}\{s\}$ such that $\sigma_{X_1}^{-1}\{s\}$ is not a singleton, and thus obtain the required surjective $h : \mathrm{Mid}(X_1) \to \mathrm{Mid}(X_2)$.

If S has no defective elements then X_1 and X_2 do not have any, and similar but simpler arguments using Theorem 5.5 complete the proof in this case. ☐

Let $\mathbb{V}$ be a finitely generated nearly regular variety. Let $\leq$ be the partial ordering of the set $\mathcal{A}$ of isomorphism types of all its critical algebras given by the requirement that $A_1 \leq A_2$ exactly when A_1 is isomorphic to a subalgebra of A_2. For critical *dp*-spaces, write $X \leq Y$ if and only if there is a surjective *dp* map $X \to Y$. Then, by Theorem 2.2(6), we have $A_1 \leq A_2$ if and only if $\mathbf{P}A_1 \leq \mathbf{P}A_2$. This observation is the starting point for the proof below.

Proof of Theorem 3.4. Let $\mathbb{V}$ be a finitely generated nearly regular variety of *dp*-algebras. Then $\mathbb{V}$ is locally finite and hence, by [AD94a], any quasivariety $\mathbb{Q} \subseteq \mathbb{V}$ is generated by its critical algebras. Critical algebras in any quasivariety form an order ideal with respect to $\leq$ and distinct quasivarieties yield distinct order ideals. Thus it suffices to show that there are only countably many order ideals of critical algebras or, equivalently, that there are only countably many order ideals of critical *dp*-spaces. We aim to prove the latter claim.

Let $\mathcal{S} \subseteq \mathbf{P}\mathbb{V}$ denote the class of all semi-rudiments which are *dp*-quotients of critical spaces in $\mathbf{P}\mathbb{V}$. Since there exist only finitely many nuclei in $\mathbb{V}$ and because every nucleus in $\mathbb{V}$ is finite, the class $\mathcal{S}$ is a finite set. For any $S \in \mathcal{S}$ consider triples (S, T_1, T_2) with $T_2 \subseteq T_1 \subseteq S$ such that

- any $t \in T_1$ is a non-defective element of $\mathrm{Mid}(S)$ comparable to another non-defective element of $\mathrm{Mid}(S)$;
- T_1 is antichain in S;
- if $|T_1| > 1$ then any $t \in T_1$ is comparable to a defective element of $\mathrm{Mid}(S)$.

Let $\mathcal{T}$ denote the set of all such triples. For any triple $(S, T_1, T_2) \in \mathcal{T}$, let $\mathcal{C}(S, T_1, T_2)$ be the set of all critical *dp*-spaces $X \in \mathbf{P}\mathbb{V}$ such that

- the semi-rudiment $S(X)$ of X is isomorphic to S; thus $\sigma_X : X \to S$ is surjective;
- $|\sigma_X^{-1}\{t\}| > 1$ for $t \in \mathrm{Mid}(S)$ if and only if $t \in T_1$;
- $\sigma_X^{-1}\{t\}$ is order disconnected if and only if $t \in T_2$.

By Theorems 5.4 and 5.5, if $\mathcal{J}$ is an order ideal of critical *dp*-spaces in $\mathbf{P}\mathbb{V}$ then $\mathcal{I} = \bigcup_{(S,T_1,T_2)} \big(\mathcal{C}(S,T_1,T_2) \cap \mathcal{J}\big) \cup (\mathcal{S} \cap \mathcal{J})$ and $\mathcal{C}(S,T_1,T_2) \cap \mathcal{J}$ is an order ideal in $\mathcal{C}(S,T_1,T_2)$ for every (S,T_1,T_2). Since the set $\mathcal{T}$ of all triples and the set $\mathcal{S}$ are finite, it suffices to prove that every poset $(\mathcal{C}(S,T_1,T_2),\leq)$ with $(S,T_1,T_2) \in \mathcal{T}$ has countably many order ideals.

Choose $(S,T_1,T_2) \in \mathcal{T}$ and a bijection $\mu : T_1 \setminus T_2 \to k$ (where, as usual, $k = \{0,1,\ldots,k-1\}$). Then define an injection $\lambda : \mathcal{C}(S,T_1,T_2) \to \mathbf{M}^k$ where $\mathbf{M}$ is the set of all natural numbers greater than 1 and $\lambda(X) = \alpha \in \mathbf{M}^k$ is given by $\alpha(i) = |\sigma_X^{-1}\{\mu^{-1}\{i\}\}|$ for all $i \in k$ and each $X \in \mathcal{C}(S,T_1,T_2)$. The element $\alpha = \lambda(X)$ of $\mathbf{M}^k$ then gives the cardinalities of the non-singleton sets $E(x) \subseteq \mathrm{Mid}(X)$ which are order connected. Order $\mathbf{M}^k$ componentwise (that is, for $\alpha,\beta \in \mathbf{M}^k$, let $\alpha \leq \beta$ exactly when $\alpha(i) \leq \beta(i)$ for all $i \in k$). By Proposition 5.6, λ is order preserving. Thus once we prove that each finite power $\mathbf{M}^k$ has countably many order ideals, the lattice of all subquasivarieties of $\mathbb{V}$ will be shown to be countable.

To this end, we consider $(\mathbf{M}^*)^k$ where $\mathbf{M}^* = \mathbf{M} \cup \{\infty\}$ extends $\mathbf{M}$ by the additional greatest element ∞, and order $(\mathbf{M}^*)^k$ componentwise. An order ideal $I \subseteq (\mathbf{M}^*)^k$ is *complete* whenever

- if $\{\alpha_j \mid j \in \mathbf{N}\} \subseteq I$ then $\sup\{\alpha_j \mid j \in \mathbf{N}\} \in I$.

Observe that the mapping from the set of all complete ideals of $(\mathbf{M}^*)^k$ to the set of all order ideals of $\mathbf{M}^k$ given by $I \mapsto I \cap (\mathbf{M}^k)$ is an order isomorphism. Thus if $(\mathbf{M}^*)^k$ has only countably many complete order ideals, then $\mathbf{M}^k$ has only countably many order ideals. For any complete ideal I of $(\mathbf{M}^*)^k$, let $\max(I)$ denote the set of all its maximal elements. Then $\max(I)$ is an antichain and, since $\mathbf{M}^*$ is a complete chain, every complete ideal I of $(\mathbf{M}^*)^k$ has the form

$$I = \{\alpha \mid \exists\beta \in \max(I),\, \alpha \leq \beta\}.$$

Now, since $(\mathbf{M}^*)^k$ is countable, it suffices to show that every antichain in $(\mathbf{M}^*)^k$ is finite.

We prove this claim by an induction over k, using an argument suggested by David Kelly. Since $\mathbf{M}^*$ is a chain, the claim obviously holds for $k = 1$. Now suppose that all antichains in $(\mathbf{M}^*)^{k-1}$ are finite, and let $A \subseteq (\mathbf{M}^*)^k$ be a countably infinite antichain. Write

$$A = \{(\beta_n, m_n) \mid n = 0,1,\ldots\},$$

where $\beta_n \in (\mathbf{M}^*)^{k-1}$ and $m_n \in \mathbf{M}^*$, and the elements of A are indexed so that $m_n \leq m_{n+1}$ for all $n \geq 0$. The elements of the set $B = \{\beta_n \mid n =$

$0, 1, \ldots\}$ with distinct indices are distinct because A is an antichain, and the antichain $M = \min(B)$ of all minimal elements of B is nonvoid and finite, by the induction hypothesis. Therefore $M \subseteq \{\beta_0, \ldots, \beta_p\}$ for some finite p and, since β_{p+1} is not minimal in B, there exists some β_r with $r \leq p$ such that $\beta_r < \beta_{p+1}$. But then $(\beta_r, m_r) < (\beta_{p+1}, m_{p+1})$ because $r < p+1$, contradicting the hypothesis that A is an antichain. Therefore every antichain $A \subseteq (\mathbf{M}^*)^k$ is finite for any finite $k \geq 1$, and the proof of Theorem 3.4 is complete. ◻

Since any antichain of $\mathbf{M}^k$ is also an antichain of $(\mathbf{M}^*)^k$, any antichain of $\mathbf{M}^k$ (and also of $\mathcal{C}(S, T_1, T_2)$) is finite. As a consequence we obtain

Corollary 5.7. *If $\mathbb{V}$ is a finitely generated nearly regular variety of dp-algebras, then every set consisting of critical algebras from $\mathbb{V}$ that have no proper injective homomorphism between them is finite.* ◻

6 Varieties of *dp*-algebras Containing a 3-semi-nucleus

In this section, we prove Theorem 3.5.

Let $\mathbb{V}$ be a finitely generated variety of *dp*-algebras containing a 3-semi-nucleus. Then $\mathbb{V}$ contains a finite algebra whose *dp*-space X is connected and has $\mathrm{Mid}(X) = \{d < y < v\}$ in which d is min-defective, while y, v are non-defective and $\mathrm{Ext}(v) \neq \mathrm{Ext}(y)$ (or $\mathbb{V}$ contains the order dual of X).

We now recall some concepts and results from [Kou90] and [KS04a].

Let $\mathbb{P}_3$ be the category whose objects are all quadruples $(\mathcal{P}, a, b, c)$, where $\mathcal{P}$ is a Priestley space and $a, b, c \in \mathrm{Ext}(\mathcal{P})$ are clopen points, and $\mathbb{P}_3$-morphisms from $(\mathcal{P}, a, b, c)$ to $(\mathcal{P}', a', b', c')$ are all Priestley maps $f : \mathcal{P} \to \mathcal{P}'$ with $f(a) = a'$, $f(b) = b'$ and $f(c) = c'$. For positive integers $k < l < m$, let $\mathbb{P}^-_{k,l,m}$ be the full subcategory of $\mathbb{P}_3$ determined by all $\mathbb{P}_3$-objects $(\mathcal{P}, a, b, c)$ in which the elements $a, b, c \in \mathrm{Min}(\mathcal{P})$ have respective distances $\mathrm{dist}(a, b) = k$, $\mathrm{dist}(a, c) = l$, $\mathrm{dist}(b, c) = m$, and

(p) $\mathrm{dist}(a, x) > k$ or $\mathrm{dist}(c, x) > m$ for every $x \in [b) \setminus \{b\}$.

Theorem 6.1 ([Kou90, KS04a]). *There exist even positive integers $k < l < m$ for which the category $\mathbb{P}^-_{k,l,m}$ is dually finite-to-finite universal.* ◻

Select the integers $k < l < m$ so that the category $\mathbb{P}^-_{k,l,m}$ is dually finite-to-finite universal and denote $\mathcal{K} = \mathbb{P}^-_{k,l,m}$.

For any $\mathcal{K}$-object $\mathcal{Y} = ((Y, \leq, \sigma), a, b, c)$ define

$$F\mathcal{Y} = ((X \setminus \{y, v\}) \cup Y \cup \{v_0, v_1\}, \leq, \sigma^+),$$

where the union is disjoint, $\{v_0, v_1\}$ is an antichain, $\leq$ is the least partial order coinciding with that of X on $X \setminus \{y, v\}$ and with that of $\mathcal{Y}$ on $Y \subseteq F\mathcal{Y}$, such that $\mathrm{Ext}_{F\mathcal{Y}}(y') = \mathrm{Ext}_X(y)$ for all $y' \in Y$ and $\mathrm{Ext}_{F\mathcal{Y}}(v_0) = \mathrm{Ext}_{F\mathcal{Y}}(v_1) = \mathrm{Ext}_X(v)$ and, additionally,

(∗) $a \leq v_0$, $d \leq b$ and $c, d \leq v_1$ in $F\mathcal{Y}$.

The topology σ^+ of $F\mathcal{Y}$ is the discrete extension of σ.

By Theorem 2.4, $\mathbf{D}F\mathcal{Y} \in \mathbb{V}$ for every $\mathcal{Y} \in \mathcal{K}$.

For any $\mathcal{K}$-morphism $f : \mathcal{Y} \to \mathcal{Y}'$, let $Ff : F\mathcal{Y} \to F\mathcal{Y}'$ be the extension of f by the identity map on the set $F\mathcal{Y} \setminus Y = X \setminus \{y\} = F\mathcal{Y}' \setminus Y'$. Then F is a well-defined functor and, in particular, $Ff(d) = d$ for any $\mathcal{K}$-morphism $f : \mathcal{Y} \to \mathcal{Y}'$.

Lemma 6.2. *If $g : F\mathcal{Y} \to F\mathcal{Y}'$ is a dp-map and $d \in \mathrm{Im}(g)$ then $g = (Ff) \circ \gamma$ for some $\mathcal{K}$-morphism $f : \mathcal{Y} \to \mathcal{Y}'$ and an automorphism γ of $F\mathcal{Y}$ moving only the elements of* $\mathrm{Ext}(F\mathcal{Y}) = \mathrm{Ext}(X)$.

Proof. Let $m \in \mathrm{Min}(F\mathcal{Y})$ be the element for which $\mathrm{Ext}(d) = \mathrm{Ext}(m)$ and let $g : F\mathcal{Y} \to F\mathcal{Y}'$ be a *dp*-map. Suppose that $d = g(z)$ for some $z \in F\mathcal{Y}$. Then $z \notin \mathrm{Ext}(F\mathcal{Y})$ and $\mathrm{Ext}(m) = g(\mathrm{Ext}(z))$. Since $\mathrm{Ext}(F\mathcal{Y}) = \mathrm{Ext}(F\mathcal{Y}') = \mathrm{Ext}(X)$ is a nucleus, the restriction g_1 of g to this set is an invertible *dp*-map, and hence $\mathrm{Ext}(z) = \mathrm{Ext}(g_1^{-1}(m))$ and $g^{-1}(m) \in \mathrm{Min}(F\mathcal{Y})$. Therefore $z \in F\mathcal{Y}$ is min-defective and, since d is the only defective element of $F\mathcal{Y}$, we have $g(d) = d$ and $g_1(m) = m$. Since $\mathrm{Mid}(F\mathcal{Y}) = \{d\} \cup Y \cup \{v_0, v_1\}$ and the sets of extremal elements of the sets Y and $\{v_0, v_1\}$ differ, we have $g(Y) \subseteq Y'$ and $g\{v_0, v_1\} \subseteq \{v_0, v_1\}$. From $d \leq v_1$ and $d \not\leq v_0$ it follows that $g(v_1) = v_1$. Now we show that $g(v_0) = v_1$ is impossible. If $g(v_0) = v_1$, then $g(a) = c'$ because $a \leq v_0$ and $(v_1] \cap Y' = \{c'\}$. Also $g(b) \geq b'$ because $[d) \cap Y = [b) \cap Y$ and $[d) \cap Y' = [b') \cap Y'$. But then $\mathrm{dist}(a, b) = k < l - 1 \leq \mathrm{dist}(c', g(b))$, a contradiction.

Therefore $g(v_0) = v_0$ and $g(v_1) = v_1$, and $g(a) = a'$ and $g(c) = c'$ then follow because $(v_0] \cap Y' = \{a'\}$ and $(v_1] \cap Y' = \{c'\}$.

By (p), for every element $x \in [b') \setminus \{b'\}$ either $\mathrm{dist}(a', x) > k$ or $\mathrm{dist}(c', x) > m$, and from this we conclude that $g(b) = b'$. Altogether, the restriction f of g to $\mathcal{Y} \subseteq F\mathcal{Y}$ maps $\mathcal{Y}$ to $\mathcal{Y}' \subseteq F\mathcal{Y}'$, and the map $f : \mathcal{Y} \to \mathcal{Y}'$ is a $\mathcal{K}$-morphism. Furthermore, $g = (Ff) \circ \gamma$ for the automorphism γ of $F\mathcal{Y}$ defined as g_1 on $\mathrm{Ext}(F\mathcal{Y})$ and as the identity map on $\mathrm{Mid}(F\mathcal{Y})$. ◻

Since $\mathcal{K}$ is finite-to-finite universal, there is a countably infinite set $\mathcal{R} = \{\mathcal{Y}_i \mid i \in \omega\}$ of finite $\mathcal{K}$-objects such that every $\mathcal{K}$-morphism $f : \mathcal{Y}_i \to \mathcal{Y}_j$ is the identity map of one of its objects. Denote $X_i = F\mathcal{Y}_i$ for $i \in \omega$. Then every X_i is a finite *dp*-space.

Lemma 6.3. *Every X_i is a critical space, and the quasivarieties $\mathbf{SPD}X_i$ generated by the algebras $\mathbf{D}X_i$ with $i \in \omega$ form an antichain.*

Proof. If $g \in \mathrm{End}\, X_i$ and $d \in \mathrm{Im}(g)$, then $g(d) = d$ and hence g is invertible, by Lemma 6.2. Thus X_i is critical, by Proposition 4.1. Suppose that $\mathbf{D}X_i$ belongs to the subquasivariety generated by $\mathbf{D}X_j$ for some $j \neq i$. Then $\mathbf{D}X_i \in \mathbf{SPD}X_j$ and hence there is a separating family $\mathcal{F} = \{f_k \mid k \in K\}$ of homomorphisms $\mathbf{D}X_i \to \mathbf{D}X_j$. By Theorem 2.2(8), the finite space X_i is the union of the images $\mathrm{Im}(\mathbf{P}f_k)$. Since $\mathcal{K}(\mathcal{Y}_i, \mathcal{Y}_j) = \emptyset$ when $j \neq i$, none of these images can contain d because of Lemma 6.2, and this contradiction proves the second claim. ◻

Theorem 3.5 immediately follows.

7 Other Varieties

As noted in Theorem 2.2(4), a *dp*-space X is dual to a regular *dp*-algebra exactly when $\mathrm{Ext}(X) = X$. In this section, we give a simple proof that a relatively small subvariety $\mathbb{R}_2$ of the variety $\mathbb{R}$ of all regular *dp*-algebras is Q-universal. Since $\mathbb{R}$ is also a variety of double Heyting algebras [Kat73], this will also prove the Q-universality of $\mathbb{R}_2$ in the case of double Heyting algebras. In contrast to *dp*-algebras, every finitely generated variety of double Heyting algebras has only finitely many critical algebras.

We begin by noting that all finite zigzags are duals of regular *dp*-algebras and of double Heyting algebras. For any odd $k \geq 3$, let Z_k denote the zigzag $0 < 1 > 2 < \ldots < k$ having k distinct comparable pairs.

Lemma 7.1. *Let $m, n \geq 3$ be odd. Then there is a dp-map $h : Z_m \to Z_n$ if and only if n divides m, and h is the identity map when $m = n$.*

Proof. For any odd zigzag Z_k and any $i \in Z_k$ we have

$$\mathrm{Ext}(i) = \begin{cases} \{0,1\} & \text{if } i = 0, \\ \{i-1, i, i+1\} & \text{if } 0 < i < k, \\ \{k-1, k\} & \text{if } i = k. \end{cases} \tag{z}$$

Let $h : Z_m \to Z_n$ be a *dp*-map, that is, let h preserve the order and $\mathrm{Ext}(h(i)) = h(\mathrm{Ext}(i))$ for every $i \in Z_m$. Since 0 is minimal in both *dp*-spaces, from (z) it follows that $h(0) = 0$ and $h(1) = 1$, and then $h(i) = i$ for any $i \leq \min\{m, n\}$. Similarly we find that $h(m) = n$, and from $h(0) = 0$ and $h(m) = n$ we obtain $m \geq n$. A simple induction using (z) then shows that $h(i) = i$ for every $i \leq n$. In particular, if $m = n$ then h is the identity mapping, and hence the second claim holds.

Suppose that $m > n$. Since $h(n) = n$, using (z) we first find $h(n+1) = n-1$, and then $h(n+j) = n-j$ for every $j \leq \min\{m, n\}$. If $m < 2n$, then $|\mathrm{Ext}(h(m))| \geq 3$ in Z_n while $|\mathrm{Ext}(m)| = 2$ in Z_m, a contradiction. Therefore $m \geq 2n$ and $h(2n) = 0$. Continuing inductively, we find that for every $k \geq 1$ and $j \in \{0, \ldots, n-1\}$ such that $m \geq kn + j$, the *dp*-map h has the form

$$h(kn + j) = \begin{cases} j & \text{if } k \text{ is even,} \\ n - j & \text{if } k \text{ is odd.} \end{cases} \tag{w}$$

Recalling that $h(m) = n$, we conclude that $m = kn$ for some (odd) $k \geq 1$.

For the converse, it is routine to verify that the map $h : Z_{kn} \to Z_n$ defined in (w) satisfies $h(\mathrm{Min}(i)) = \mathrm{Min}(h(i))$ and $h(\mathrm{Max}(i)) = \mathrm{Max}(h(i))$ for every $i = 0, \ldots, kn$. ◻

In [Lee70], K. B. Lee showed that the varieties of distributive p-algebras form a chain

$$\mathbb{T} \subset \mathbb{B}_0 \subset \mathbb{B}_1 \subset \ldots \subset \mathbb{B}_n \subset \ldots \subset \mathbb{B}_\infty$$

in which $\mathbb{T}$ is the trivial class, $\mathbb{B}_0$ consists of all Boolean algebras and, for any $n \geq 1$, the variety $\mathbb{B}_n$ is formed by all p-algebras satisfying the identity

$$(x_1 \wedge \ldots \wedge x_n)^* \vee (x_1^* \wedge \ldots \wedge x_n)^* \vee \ldots \vee (x_1 \wedge \ldots \wedge x_n^*)^* = 1.$$

The variety $\mathbb{B}_n$ with $n \geq 1$ is generated by the p-algebra $B_n = 2^n + 1$ obtained from the Boolean algebra 2^n by adding 1 as the new unit element. The Priestley space $X_n = \mathrm{Ext}(X_n)$ of B_n then satisfies $|\,\mathrm{Max}(X_n)| = 1$ and $|\,\mathrm{Min}(X_n)| = n$, and hence a p-algebra A belongs to $\mathbb{B}_n$ exactly when its Priestley space $X = \mathbf{P}A$ satisfies $|\,\mathrm{Min}(x)| \leq n$ for every $x \in \mathrm{Max}(X)$.

Let $\mathbb{R}_n$ denote the variety of dp-algebras satisfying the above identity along with its order dual

$$(x_1 \vee \ldots \vee x_n)^+ \wedge (x_1^+ \vee \ldots \vee x_n)^+ \wedge \ldots \wedge (x_1 \vee \ldots \vee x_n^+)^+ = 0.$$

It is then clear that every zigzag belongs to the subvariety $\mathbb{R}_2$ of $\mathbb{R}$. By Lemma 7.1, the zigzags whose lengths are pairwise distinct odd primes then form an infinite set of dp-spaces of finite hereditarily simple algebras not embeddable into one another, and Proposition 2.1 in Dziobiak [Dzi85a] then yields

Corollary 7.2. *The variety $\mathbb{R}_2$ of dp-algebras is Q-universal.* □

We recall that a *double Heyting algebra* $A = (L, \vee, \wedge, \rightarrow, \leftarrow, 0, 1)$ is an algebra of type $(2, 2, 2, 2, 0, 0)$ such that $(L, \vee, \wedge, 0, 1)$ is a distributive $(0, 1)$-lattice and $\rightarrow$ and $\leftarrow$ are binary operations, respectively determined by the requirement that $t \leq (x \rightarrow y)$ exactly when $t \wedge x \leq y$, while $t \geq (x \leftarrow y)$ exactly when $t \vee x \geq y$. Since regular dp-algebras are also double Heyting algebras [Kat73], and because $x^* = x \rightarrow 0$ and $x^+ = x \leftarrow 1$, from [Lee70] it follows that $\mathbb{R}_2$ is a variety of double Heyting algebras. We thus have

Corollary 7.3. *The variety $\mathbb{R}_2$ of double Heyting algebras is Q-universal.* □

Remark. Define $a^{0(+*)} = a$ and $a^{(k+1)(+*)} = a^{k(+*)+*}$ for $k \geq 0$, and recall that a variety $\mathbb{V}$ of dp-algebras is of *finite range* $n \geq 0$ if $a^{(n+1)(+*)} = a^{n(+*)}$ whenever $a \in A \in \mathbb{V}$. When combined, W. Dziobiak's results [Dzi85a, Dzi85b] show the existence of a Q-universal variety $\mathbb{V}_4 \subseteq \mathbb{R}$ of range four, and that $\mathbb{R}_2 \cap \mathbb{V}_4$ is generated by the 3-element chain. The variety $\mathbb{R}_2 \cap \mathbb{V}_4$ is not Q-universal, see [KS04b], for instance.

The easy claim below contrasts our results about finitely generated varieties of dp-algebras.

Observation 7.4. *Any finitely generated variety of double Heyting algebras has only finitely many critical algebras, and hence it is not Q-universal.*

Proof. Let $\mathbb{V}$ be a finitely generated variety of double Heyting algebras. Every finite algebra $A \in \mathbb{V}$ is isomorphic to a direct product $A_1 \times \ldots \times A_n$ of subdirectly irreducible algebras $A_i \in \mathbb{V}$ (all of which are simple), see [KS91, KS94], for instance. If A is critical, then no two distinct factors A_i and A_j can be isomorphic. Since $\mathbb{V}$ has only finitely many non-isomorphic subdirectly irreducible algebras, it has only finitely many non-isomorphic critical algebras. ◻

Remark. It may be of interest to recall from [KS90] that $\mathbb{R}_3$ is categorically universal, while $\mathbb{R}$ is not finite-to-finite universal in either case, see [KS98a] and [KS98b], respectively.

References

[AHS90] J. Adámek, H. Herrlich and G. E. Strecker, *Abstract and Concrete Categories,* Wiley, New York, 1990.

[AD94a] M. E. Adams and W. Dziobiak, Lattices of quasivarieties of 3-element algebras, *J. of Algebra* **166** (1994), 181–210.

[AD94b] M. E. Adams and W. Dziobiak, Q-universal quasivarieties of algebras, *Proc. Amer. Math. Soc.* **120** (1994), 1053–1059.

[AD01] M. E. Adams and W. Dziobiak, Finite-to-finite universal quasivarieties are Q-universal, *Alg. Universalis* **46** (2001), 253–283.

[Dav78] B. A. Davey, Subdirectly irreducible distributive double p-algebras, *Algebra Universalis* **8** (1978), 73–88.

[DK05] M. Demlová and V. Koubek, Weak alg-universality and Q-universality in semigroup quasivarieties, *Comment. Math. Univ. Carolin.* **46** (2005), 257–279.

[Dzi85a] W. Dziobiak, On subquasivariety lattices of some varieties related with distributive p-algebras, *Algebra Universalis* **21** (1985), 205–214.

[Dzi85b] W. Dziobiak, The subvariety lattice of the variety of distributive double p-algebras, *Bull. Austral. Math. Soc.* **31** (1985), 377–387.

[Dzi86] W. Dziobiak, On lattice identities satisfied in subquasivariety lattices of varieties of modular lattices, *Algebra Universalis* **22** (1986), 205–214.

[Gor98] V. A. Gorbunov, *Algebraic theory of quasivarieties,* Plenum, New York, 1998.

[Kat73] T. Katriňák, The structure of distributive double p-algebras, regularity and congruences, *Algebra Universalis* **21** (1973), 238–246.

[Kou85] V. Koubek, Infinite image homomorphisms of distributive bounded lattices, in *Proc. Colloquia Math. Soc. János Bolyai, 43. Lectures in Universal Algebra, Szeged 1983* North Holland, 1985, 241–281.

[Kou90] V. Koubek, Finite-to-finite universal varieties of distributive double p-algebras, *Comment. Math. Univ. Carolin.* **31** (1990), 67–83.

[KS90] V. Koubek and J. Sichler, Categorical universality of regular distributive double p-algebras, *Glasgow Math. J.* **32** (1990), 243–256.

[KS91] V. Koubek and J. Sichler, Priestley duals of products, *Cahiers Topologie Gèom. Diff. Catégoriques* **32** (1991), 243–256.

[KS94] V. Koubek and J. Sichler, Universal finitely generated varieties of double p-algebras, *Cahiers Topologie Géom. Diff. Catégoriques* **35** (1994), 139–164.

[KS98a] V. Koubek and J. Sichler, Equimorphy in varieties of double Heyting algebras, *Coll. Math.* **77** (1998), 41–58.

[KS98b] V. Koubek and J. Sichler, Equimorphy in varieties of distributive double p-algebras, *Czech. Math. J.* **48** (1998), 473–544.

[KS04a] V. Koubek and J. Sichler, Finitely generated varieties of distributive double p-algebras universal modulo a group, *Algebra Universalis* **51** (2004), 35–79.

[KS04b] V. Koubek and J. Sichler, On relative universality and Q-universality, *Studia Logica* **78** (2004), 279–291.

[KS06$^+$] V. Koubek and J. Sichler, On subquasivarieties of finitely generated varieties of distributive double p-algebras, to appear in *Contrib. Gen. Algebra* **17**.

[Lee70] K. B. Lee, Equational classes of distributive pseudocomplemented lattices, *Canad. J. Math.* **22** (1970), 881–891.

[Pri70] H. A. Priestley, Representation of distributive lattices by means of ordered Stone spaces, *Bull. London Math. Soc.* **2** (1970), 186–190.

[Pri84] H. A. Priestley, Ordered sets and duality for distributive lattices, *Ann. Discrete Math.* **23** (1984), 36–60.

[Sap85] M. V. Sapir, The lattice of quasivarieties of semigroups, *Algebra Universalis* **21** (1985), 172–180.

The F-triangle of the Generalised Cluster Complex

Christian Krattenthaler*

Institut Camille Jordan, Université Claude Bernard Lyon-I, 21, avenue Claude Bernard, F-69622 Villeurbanne Cedex, France**
http://igd.univ-lyon1.fr/~kratt

Summary. The F-triangle is a refined face count for the generalised cluster complex of Fomin and Reading. We compute the F-triangle explicitly for all irreducible finite root systems. Furthermore, we use these results to partially prove the "$F = M$ Conjecture" of Armstrong which predicts a surprising relation between the F-triangle and the Möbius function of his m-divisible partition poset associated to a finite root system.

AMS Subject Classification. Primary 05E15; Secondary 05A05 05A10 05A15 05A19 06A07 20F55 33C05.
Keywords. root systems, generalised cluster complex, face numbers, F-triangle, m-divisible non-crossing partitions, Möbius function, M-triangle, Chu–Vandermonde summation.

1 Introduction

Fomin and Zelevinsky created a new exciting research field when they invented *cluster algebras* in [FZ02]. The classification of cluster algebras *of finite type* from [FZ03a] says that there is a one-to-one correspondence between finite-type cluster algebras and finite root systems. Furthermore, for each finite root system Φ, Fomin and Zelevinsky [FZ03b] defined a simplicial complex corresponding to the associated cluster algebra, the *cluster complex* $\Delta(\Phi)$. This is a simplicial complex on a subset of the set of roots Φ. As they showed, this complex has many remarkable properties. In particular, the number of facets is given by the *Catalan number for the root system* Φ, and, moreover, all the face numbers are given by elegant product formulae. Further remarkable (originally, conjectural) properties have been discovered by Chapoton in

* Research partially supported by EC's IHRP Programme, grant HPRN-CT-2001-00272, "Algebraic Combinatorics in Europe."

** Current address: Fakultät für Mathematik, Universität Wien, Nordbergstraße 15, A-1090 Vienna, Austria.

[Cha04]. In this paper, he refines the face enumeration to, what he calls, the "F-triangle." He computed the F-triangle for all types (and revealed his findings partially in [Cha04]) and observed a surprising relationship (see [Cha04, Conjecture 1]) between the F-triangle and the Möbius function of the non-crossing partition lattice $NC(\Phi)$ associated to Φ, the latter being due to Bessis [Bes03] and Brady and Watt [BW02]. This relationship, to which we shall refer in the sequel as the "$F = M$ Conjecture," has been recently proved by Athanasiadis [Ath06$^+$]. For further fascinating properties of the F-triangle see [Cha04].

The subject of this paper is "m-generalisations" of the cluster complex $\Delta(\Phi)$ and of the non-crossing partition lattice $NC(\Phi)$. More precisely, in [FR05], Fomin and Reading introduce the generalised cluster complex $\Delta^m(\Phi)$, where m is some non-negative integer. This is, again, a simplicial complex, now on *coloured* roots, and for $m = 1$ it reduces to the (ordinary) cluster complex $\Delta(\Phi)$. As they show, this generalised complex has again remarkable properties. In particular, the number of facets is given by the *Fuss–Catalan number for the root system* Φ, and, moreover, again, all the face numbers are given by elegant product formulae.

Going one step further, Sergey Fomin suggested to the author to investigate the "Chapoton-like" refinement of the face numbers of the generalised cluster complex $\Delta^m(\Phi)$, that is to say, to study the "F-triangle" for $\Delta^m(\Phi)$ (see Section 2 for the definition). This is what we do in this paper. We compute the F-triangle of $\Delta^m(\Phi)$ for all types of irreducible root systems Φ, see Sections 4–7. We do this case-by-case. While a case-independent formula would certainly be desirable, certain features of our results (in particular, the appearance of the Kronecker delta in the formula in Theorem FD in Section 6 for type D_n) make it highly unlikely that such a case-independent formula exists. As an aside, we draw the reader's attention to the unexpected outcome of our results that the refined face numbers are all polynomials in m *with non-negative coefficients*, a phenomenon for which we have no intrinsic explanation.

One may then ask if there is also an "$F = M$ Conjecture" in this generalised context. This would, first of all, require an "m-extension" of the non-crossing partition lattice. Indeed, Armstrong [Arm06$^+$] has recently introduced the "m-divisible non-crossing partition poset" $NC^m(\Phi)$, generalising an earlier construction of Edelman [Ede81] in type A_n. He shows that this poset has also remarkable properties, resembling those of the non-crossing partition lattices. Moreover, he observed that there is also a rather straightforward extension of the $F = M$ Conjecture relating the F-triangle of the generalised cluster complex $\Delta^m(\Phi)$ to the Möbius function of the corresponding m-divisible non-crossing partitions poset $NC^m(\Phi)$. We reproduce this conjecture in Section 8. (We refer the reader to [Arm06$^+$, Sec. 4] and [Tza06$^+$] for further fascinating properties of the F-triangle of $\Delta^m(\Phi)$.)

With the explicit formulae for the F-triangle in hand, we are able to prove this "m-version" of the $F = M$ Conjecture in types A_n and B_n, for the dihedral root systems $I_2(a)$, for the hyperbolic root systems H_3 and H_4, and for F_4 and E_6, see Sections 9, 10, 13–17. In types A_n and B_n, the proofs depend crucially on results about rank selected chain enumeration in $NC^m(A_n)$ and $NC^m(B_n)$ due to Edelman [Ede81] and Armstrong [Arm06+], respectively. Moreover, in Section 11, we provide a calculation in type D_n which will prove the conjecture also for this type once the corresponding rank selected chain enumeration result analogous to the ones by Edelman and Armstrong is available for $NC^m(D_n)$. In view of the results of Athanasiadis and Reiner [AR04] on rank selected chain enumeration in $NC^1(D_n) = NC(D_n)$, this argument does accomplish the proof for $m = 1$. As we explain in Section 12, the verification of the (generalised) $F = M$ Conjecture in the exceptional types is a routine task which can, in principle, be carried out on a computer. To do this in practice for the root systems E_7 or E_8, say, may however require additional simplifications of the proposed procedure.[1]

In the final Section 18, we prove Armstrong's conjecture [Arm06+, Sec. 4] on the form of, what he calls, the *dual F-triangle* in a case-by-case fashion. For the exceptional root systems this is just a routine calculation, while for A_n, B_n and D_n this requires only the Chu–Vandermonde summation formula.

We conclude the introduction by saying a few words how the F-triangles for $\Delta^m(\Phi)$ are found and the corresponding results are proved in this paper. The main tool in [Cha04] for finding F-triangles for the (ordinary) cluster complex $\Delta(\Phi)$ consists in two recurrence formulae (see [Cha04, Prop. 3]). These recurrences carry over verbatim to the generalised cluster complex $\Delta^m(\Phi)$, see Proposition F in Section 2. Indeed, in the exceptional types, the formula for the F-triangle for $\Delta^m(\Phi)$ can be found in a routine fashion by using these recurrences, see Section 7. In types A_n, B_n, and D_n, however, the recurrences can only be used to compute the F-triangle for the corresponding generalised cluster complex for *specific* n. By doing this for sufficiently many n, we first worked out *guesses* for the F-triangle for *generic* n. (In types A_n and B_n, this has also been done independently by Tzanaki [Tza06+].) Subsequently, one tries to verify these guesses by checking the recurrences. As it turns out, this requires multivariate summation formulae due to Carlitz [Car77], which we restate here in Section 3 for the convenience of the reader. An interesting detail is the fact that Chapoton's proofs in [Cha04] for the F-triangle for $\Delta(A_n)$ and $\Delta(B_n)$, respectively, which also use Carlitz's summation formulae, do *not* extend to $\Delta^m(A_n)$ and $\Delta^m(B_n)$, for the following reason. In order to do the above described verification using the recurrences, he has to evaluate a triple sum. He does this by first simplifying one sum by means of the Chu–Vandermonde summation, and by using subsequently one of Carlitz's summation formulae to evaluate the remaining double sum. However, if $m \neq 1$, the Chu–Vandermonde summation is not applicable to the triple sum

[1] Using some additional ideas, this has been carried through in *"The M-triangle of generalised non-crossing partitions for the types E_7 and E_8"* (preprint; arXiv:math.CO/0601676).

that we encounter at the start. Remarkably, it is possible to apply Carlitz's summation formula *directly*, in a *different* way than in [Cha04]. The use of the Chu–Vandermonde summation is then not necessary anymore.

2 Preliminaries

Let Φ be a finite root system of rank n. (We refer the reader to [Hum90] for all root system terminology.) For a non-negative integer m, the generalised cluster complex $\Delta^m(\Phi)$ is a certain simplicial complex on a certain set of "coloured" roots, the roots being from Φ. The precise definition will not be important here, we refer the reader to [FR05, Sec. 2]. The only fact which is important here is that some of the coloured roots can be positive, others negative. Let $f_{k,l}(\Phi,m)$ denote the number of faces of $\Delta^m(\Phi)$ which contain exactly k positive and l negative coloured roots. Define the F-triangle of $\Delta^m(\Phi)$, denoted by $F_\Phi^m(x,y)$, as the two-variable polynomial

$$F_\Phi^m(x,y) = \sum_{k,l\ge 0} f_{k,l}(\Phi,m)\, x^k y^l. \tag{1}$$

It is called "triangle" because all faces have cardinality at most n and, thus, in the summation in (1) we can restrict the summation indices to the triangle $k+l\le n$, $k,l\ge 0$.

Then, in this generalised context, the arguments from [Cha04, Prop. 3] carry over verbatim to prove the following properties of the F-triangle of $\Delta^m(\Phi)$.

Proposition F. *The F-triangle $F_\Phi^m(x,y)$ satisfies the following three properties:*

1. *If Φ and Φ' are two root systems, then*

$$F_{\Phi\times\Phi'}^m(x,y) = F_\Phi^m(x,y) F_{\Phi'}^m(x,y), \tag{2}$$

where $\Phi\times\Phi'$ denotes the orthogonal product of the two root systems.

2. *If $\Phi = \Phi(S)$ is an irreducible root system with simple roots S, then*

$$\frac{\partial}{\partial y} F_{\Phi(S)}^m(x,y) = \sum_{\alpha\in S} F_{\Phi(S\backslash\{\alpha\})}^m(x,y), \tag{3}$$

where $\Phi(S\backslash\{\alpha\})$ denotes the root system generated by the simple roots $S\backslash\{\alpha\}$.

3. *The specialisation $x=y$ is given by*

$$F_\Phi^m(x,x) = \sum_{k\ge 0} f_k(\Phi,m)\, x^k, \tag{4}$$

where the coefficients $f_k(\Phi, m)$ *are the face numbers of the cluster complex* $\Delta^m(\Phi)$*, summarised in* [FR05, Theorem 7.5] *for the irreducible root systems.*

We remark that an equivalent statement of (3) is

$$l \cdot f_{k,l}(\Phi(S), m) = \sum_{\alpha \in S} f_{k,l-1}(\Phi(S \backslash \{\alpha\}), m), \quad k, l \geq 0. \tag{5}$$

Moreover, in view of the multiplicativity property (2), it suffices to compute the F-triangle for the irreducible root systems, which we do in Sections 4–7.

3 Carlitz's Summation Formulae

Crucial in the proofs of our claims for the F-triangle in types A_n, B_n and D_n are the following two double sum evaluations due to Carlitz [Car77]. (He has in fact extensions for any number of summations, see [Car77, Sec. 6].) Let

$$A_{k,n}(\alpha, \beta) = \frac{bk\alpha + cn\beta + \alpha\beta}{(ak + cn + \alpha)(bk + dn + \beta)} \binom{ak + cn + \alpha}{k} \binom{bk + dn + \beta}{n}.$$

Here, and in the sequel, for integers N and K the binomial coefficient $\binom{N}{K}$ is understood according to the definition

$$\binom{N}{K} = \begin{cases} \frac{N(N-1)\cdots(N-K+1)}{K!} & \text{if } K \geq 0, \\ 0 & \text{if } K < 0. \end{cases} \tag{6}$$

Then (see [Car77, (5.14)]),

$$\sum_{k_1, n_1 \geq 0} A_{k_1, n_1}(\alpha, \beta) A_{k-k_1, n-n_1}(\alpha', \beta') = A_{k,n}(\alpha + \alpha', \beta + \beta'). \tag{7}$$

Furthermore (see [Car77, (5.15); the minus sign in front of cn must be replaced by a plus sign there]),

$$\sum_{k_1, n_1 \geq 0} \binom{ak_1 + cn_1 + \alpha - 1}{k_1} \binom{bk_1 + dn_1 + \beta - 1}{n_1} A_{k-k_1, n-n_1}(\alpha', \beta')$$
$$= \binom{ak + cn + \alpha + \alpha' - 1}{k} \binom{bk + dn + \beta + \beta' - 1}{n}. \tag{8}$$

4 The F-triangle for A_n

The theorem below gives an explicit expression for the refined face numbers $f_{k,l}(A_n, m)$, and, thus, of the F-triangle in type A_n.

Theorem FA. *For $n \geq 1$, the face numbers $f_{k,l}(A_n, m)$ are given by*

$$f_{k,l}(A_n, m) = \frac{l+1}{k+l+1} \binom{n}{k+l} \binom{m(n+1)+k-1}{k}.$$

Proof. In view of Proposition F and (5) in Section 2, in order to prove this claim we have to show

$$l \cdot f_{k,l}(A_n, m) = \sum_{\substack{n_1+n_2=n-1\\ k_1+k_2=k\\ l_1+l_2=l-1}} f_{k_1,l_1}(A_{n_1}, m) \cdot f_{k_2,l_2}(A_{n_2}, m) \tag{9}$$

and

$$\sum_{k_1+l_1=k} f_{k_1,l_1}(A_n, m) = \frac{1}{k+1} \binom{n}{k} \binom{m(n+1)+k+1}{k}. \tag{10}$$

The triple sum on the right-hand side of (9) is

$$\sum_{k_1,l_1,n_1 \geq 0} \frac{l_1+1}{k_1+l_1+1} \binom{n_1}{k_1+l_1} \binom{m(n_1+1)+k_1-1}{k_1} \\ \cdot \frac{l-l_1}{k-k_1+l-l_1} \binom{n-n_1-1}{k-k_1+l-l_1-1} \binom{m(n-n_1)+k-k_1-1}{k-k_1}.$$

We replace n_1 by $n_1+k_1+l_1$ and rewrite the resulting expression in the form

$$\sum_{k_1,l_1,n_1 \geq 0} \frac{m(l_1+1)}{m(k_1+l_1+n_1+1)+k_1} \binom{n_1+k_1+l_1+1}{n_1} \\ \cdot \binom{m(n_1+k_1+l_1+1)+k_1}{k_1} \frac{m(l-l_1)}{m(n-n_1-k_1-l_1)+k-k_1} \\ \cdot \binom{n-n_1-k_1-l_1}{n-k-l-n_1} \binom{m(n-n_1-k_1-l_1)+k-k_1}{k-k_1}. \tag{11}$$

Forgetting the sum over l_1, this is now exactly in the form of the left-hand side of (7) with n replaced by $n-k-l$, $a = m+1$, $c = m$, $\alpha = m(l_1+1)$, $\alpha' = m(l-l_1)$, $b = d = 1$, $\beta = l_1+1$, and $\beta' = l-l_1$. Substituting the right-hand side, we obtain

$$\sum_{l_1=0}^{l-1} A_{k,n-k-l}(m(l+1), l+1)$$

for the sum in (11), or, equivalently,

$$l \cdot \frac{m(l+1)}{m(n+1)+k} \binom{m(n+1)+k}{k} \binom{n+1}{n-k-l},$$

which is indeed equal to $l \cdot f_{k,l}(A_n, m)$. This proves (9).

Next we compute the sum on the left-hand side of (10),

$$\sum_{k_1=0}^{k} \frac{k-k_1+1}{k+1} \binom{n}{k} \binom{m(n+1)+k_1-1}{k_1}$$
$$= \sum_{k_1=0}^{k} \binom{n}{k} \binom{m(n+1)+k_1-1}{k_1} - \sum_{k_1=0}^{k} \frac{m(n+1)}{k+1} \binom{n}{k} \binom{m(n+1)+k_1-1}{k_1-1}$$
$$= \binom{n}{k} \binom{m(n+1)+k}{m(n+1)} - \frac{m(n+1)}{k+1} \binom{n}{k} \binom{m(n+1)+k}{m(n+1)+1}$$
$$= \frac{1}{k+1} \binom{n}{k} \binom{m(n+1)+k+1}{k},$$

the simplification of summations being due to the Chu–Vandermonde summation (see [GKP89, Sec. 5.1, (5.27)] or [Sla66, (1.7.7), Appendix (III.4)]). This completes the proof. □

5 The F-triangle for B_n

The theorem below gives an explicit expression for the refined face numbers $f_{k,l}(B_n, m)$, and, thus, of the F-triangle in type B_n.

Theorem FB. *For $n \geq 1$, the face numbers $f_{k,l}(B_n, m)$ are given by*

$$f_{k,l}(B_n, m) = \binom{n}{k+l} \binom{mn+k-1}{k}.$$

Here we identify B_1 with A_1.

Proof. By inspection, the formula for $f_{k,l}(B_1, m)$ given in the theorem agrees with the formula for $f_{k,l}(A_1, m)$ in Theorem FA. Hence, in view of Proposition F and (5) in Section 2, in order to prove this claim we have to show

$$l \cdot f_{k,l}(B_n, m) = \sum_{\substack{n_1+n_2=n-1 \\ k_1+k_2=k \\ l_1+l_2=l-1}} f_{k_1,l_1}(B_{n_1}, m) \cdot f_{k_2,l_2}(A_{n_2}, m) \tag{12}$$

and

$$\sum_{k_1+l_1=k} f_{k_1,l_1}(B_n, m) = \binom{n}{k} \binom{mn+k}{k}. \tag{13}$$

The triple sum on the right-hand side of (12) is

$$\sum_{k_1,l_1,n_1 \geq 0} \binom{n_1}{k_1+l_1} \binom{mn_1+k_1-1}{k_1}$$
$$\cdot \frac{l-l_1}{k-k_1+l-l_1} \binom{n-n_1-1}{k-k_1+l-l_1-1} \binom{m(n-n_1)+k-k_1-1}{k-k_1}.$$

We replace n_1 by $n_1+k_1+l_1$ and rewrite the resulting expression in the form

$$\sum_{k_1,l_1,n_1\ge 0}\binom{n_1+k_1+l_1}{n_1}\binom{m(n_1+k_1+l_1)+k_1-1}{k_1}\\ \cdot\frac{m(l-l_1)}{m(n-n_1-k_1-l_1)+k-k_1}\binom{n-n_1-k_1-l_1}{n-k-l-n_1}\\ \cdot\binom{m(n-n_1-k_1-l_1)+k-k_1}{k-k_1}. \tag{14}$$

Forgetting the sum over l_1, this is now in the form of the left-hand side of (8) with n replaced by $n-k-l$, $a=m+1$, $c=m$, $\alpha=ml_1$, $\alpha'=m(l-l_1)$, $b=d=1$, $\beta=l_1+1$, $\beta'=l-l_1$. Substituting the right-hand side, we obtain

$$\sum_{l_1=0}^{l-1}\binom{mn+k-1}{k}\binom{n}{n-k-l}$$

for the sum in (14), or, equivalently,

$$l\cdot\binom{n}{k+l}\binom{mn+k-1}{k},$$

which is indeed equal to $l\cdot f_{k,l}(B_n,m)$. This proves (12).

Next we compute the sum on the left-hand side of (13),

$$\sum_{k_1=0}^{k}\binom{n}{k}\binom{mn+k_1-1}{k_1}=\binom{n}{k}\binom{mn+k}{k},$$

the simplification of summation being due to the Chu–Vandermonde summation. This completes the proof. □

6 The F-triangle for D_n

The theorem below gives an explicit expression for the refined face numbers $f_{k,l}(D_n,m)$, and, thus, of the F-triangle in type D_n.

Theorem FD. *For $n\ge 2$, the face numbers $f_{k,l}(D_n,m)$ are given by*

$$f_{k,l}(D_n,m)=\binom{n}{k+l}\binom{m(n-1)+k-1}{k}+\\ m\binom{n-1}{k+l-1}\binom{m(n-1)+k-2}{k-1}-\delta_{l,0}\frac{1}{n-1}\binom{n-1}{k-1}\binom{m(n-1)+k-1}{k},$$

where $\delta_{l,0}$ is the Kronecker delta, that is, it is equal to 1 if $l=0$, and it is equal to 0 otherwise. Here we identify D_2 with A_1^2, and we identify D_3 with A_3.

Proof. By inspection, for $n=2$ the formula for $f_{k,l}(D_2,m)$ given in the theorem yields for the F-triangle

$$F^m_{D_2}(x,y)=\sum_{k,l\ge 0} f_{k,l}(D_2,m)\,x^k y^l$$
$$=m^2x^2+2mxy+y^2+2mx+2y+1=(mx+y+1)^2,$$

which, according to Theorem FA, is indeed the F-triangle of A_1^2. Furthermore, again by inspection, the formula for $f_{k,l}(B_3,m)$ given in the theorem agrees with the formula for $f_{k,l}(A_3,m)$ in Theorem FA. Hence, in view of Proposition F and (5) in Section 2, in order to prove this claim we have to show

$$l\cdot f_{k,l}(D_n,m)=\sum_{\substack{n_1+n_2=n-1,\ n_1\ge 2\\ k_1+k_2=k\\ l_1+l_2=l-1}} f_{k_1,l_1}(D_{n_1},m)\cdot f_{k_2,l_2}(A_{n_2},m)+2\cdot f_{k,l-1}(A_{n-1},m) \tag{15}$$

and

$$\sum_{k_1+l_1=k} f_{k_1,l_1}(D_n,m)=\binom{n}{k}\binom{m(n-1)+k}{k}+\binom{n-2}{k-2}\binom{m(n-1)+k-1}{k}. \tag{16}$$

We start with the proof of (15). Clearly, it suffices to consider the case $l\ge 1$ because (15) is trivially true for $l=0$. We shall therefore assume $l\ge 1$ from now on.

Using the rewriting

$$\frac{l+1}{k+l+1}\binom{n}{k+l}\binom{m(n+1)+k-1}{k}$$
$$=\frac{m(l+1)}{m(n+1)+k}\binom{n+1}{n-k-l}\binom{m(n+1)+k}{k}$$

of the defining expression for $f_{k,l}(A_n,m)$ in Theorem FA, the expression on the right-hand side of (15) is

$$\sum_{l_1=0}^{l-1}\sum_{k_1=0}^{k}\sum_{n_1=2}^{n-1}\binom{n_1}{k_1+l_1}\binom{m(n_1-1)+k_1-1}{k_1}$$
$$\cdot\frac{m(l-l_1)}{m(n-n_1)+k-k_1}\binom{n-n_1}{n-n_1-k+k_1-l+l_1}\binom{m(n-n_1)+k-k_1}{k-k_1} \tag{17a}$$

$$+\,m\sum_{l_1=0}^{l-1}\sum_{k_1=0}^{k}\sum_{n_1=2}^{n-1}\binom{n_1-1}{k_1+l_1-1}\binom{m(n_1-1)+k_1-2}{k_1-1} \tag{17b}$$
$$\cdot\frac{m(l-l_1)}{m(n-n_1)+k-k_1}\binom{n-n_1}{n-n_1-k+k_1-l+l_1}\binom{m(n-n_1)+k-k_1}{k-k_1}$$

$$-\sum_{k_1=0}^{k}\sum_{n_1=2}^{n-1}\frac{1}{n_1-1}\binom{n_1-1}{k_1-1}\binom{m(n_1-1)+k_1-1}{k_1}$$
$$\cdot\frac{ml}{m(n-n_1)+k-k_1}\binom{n-n_1}{n-n_1-k+k_1-l}\binom{m(n-n_1)+k-k_1}{k-k_1} \tag{17c}$$
$$+\frac{2l}{k+l}\binom{n-1}{k+l-1}\binom{mn+k-1}{k}. \tag{17d}$$

We treat the three sums in (17) separately. We begin with the sum (17a). We extend the sum to all $n_1\ge 0$. In order to do so, we must subtract the terms with $n_1=0$ and $n_1=1$. If $n_1=0$, the summand is only non-zero for $k_1=l_1=0$ because of the presence of the binomial $\binom{n_1}{k_1+l_1}$. If $n_1=1$, then, because of the presence of the binomial coefficient $\binom{m(n_1-1)+k_1-1}{k_1}=\binom{k_1-1}{k_1}$, the summand can be non-zero only if $k_1=0$. On the other hand, in that case, the summand can be only non-zero for $l_1=0$ and $l_1=1$, again because of the presence of the binomial $\binom{n_1}{k_1+l_1}=\binom{1}{l_1}$. In summary, the sum (17a) is equal to

$$\sum_{l_1=0}^{l-1}\sum_{k_1=0}^{k}\sum_{n_1=0}^{n-1}\Bigg(\binom{n_1}{k_1+l_1}\binom{m(n_1-1)+k_1-1}{k_1}$$
$$\cdot\frac{m(l-l_1)}{m(n-n_1)+k-k_1}\binom{n-n_1}{n-n_1-k+k_1-l+l_1}\binom{m(n-n_1)+k-k_1}{k-k_1}\Bigg)$$
$$-\frac{l}{k+l}\binom{n-1}{k+l-1}\binom{mn+k-1}{k}-\frac{l}{k+l}\binom{n-2}{k+l-1}\binom{m(n-1)+k-1}{k}$$
$$-\frac{l-1}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-1}{k},$$

where the second-to-last term corresponds to the summand for $n_1=k_1=l_1=0$, the next-to-last term corresponds to the summand for $n_1=1$, $k_1=0$, $l_1=0$, and the last term corresponds to the summand for $n_1=1$, $k_1=0$, $l_1=1$. In the sum over n_1,k_1,l_1, we replace n_1 by $n_1+k_1+l_1$. Forgetting the sum over l_1, we see that it is then in the form of the left-hand side of (8) with n replaced by $n-k-l$, $a=m+1$, $c=m$, $\alpha=m(l_1-1)$, $\alpha'=m(l-l_1)$, $b=d=1$, $\beta=l_1+1$, $\beta'=l-l_1$. Hence, if we substitute the right-hand side, the expression simplifies to

$$\sum_{l_1=0}^{l-1}\binom{n}{k+l}\binom{m(n-1)+k-1}{k}-\frac{l}{k+l}\binom{n-1}{k+l-1}\binom{mn+k-1}{k}$$
$$-\frac{l}{k+l}\binom{n-2}{k+l-1}\binom{m(n-1)+k-1}{k}$$
$$-\frac{l-1}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-1}{k}. \tag{18}$$

Clearly, the sum over l_1 sums the same summand for each l_1, so that the result is that summand multiplied by l.

We next turn our attention to the sum (17b). The first observation is that for $k_1 = 0$ the summand vanishes because of the presence of the binomial coefficient $\binom{m(n_1-1)+k_1-2}{k_1-1}$. We therefore replace k_1 by k_1+1 to obtain

$$m\sum_{l_1=0}^{l-1}\sum_{k_1=0}^{k-1}\sum_{n_1=2}^{n-1}\binom{n_1-1}{k_1+l_1}\binom{m(n_1-1)+k_1-1}{k_1}\frac{m(l-l_1)}{m(n-n_1)+k-k_1-1}\\ \cdot\binom{n-n_1}{n-n_1-k+k_1-l+l_1+1}\binom{m(n-n_1)+k-k_1-1}{k-k_1-1}.$$

This time we extend the sum to $n_1 \ge 1$. In order to do so, we must subtract the terms with $n_1 = 1$. In the latter case, because of the presence of the binomial coefficient $\binom{n_1-1}{k_1+l_1}$, the summand will vanish except if $k_1 = l_1 = 0$. Thus, we obtain

$$m\sum_{l_1=0}^{l-1}\sum_{k_1=0}^{k-1}\sum_{n_1=1}^{n-1}\left(\binom{n_1-1}{k_1+l_1}\binom{m(n_1-1)+k_1-1}{k_1}\frac{m(l-l_1)}{m(n-n_1)+k-k_1-1}\right.\\ \left.\cdot\binom{n-n_1}{n-n_1-k+k_1-l+l_1+1}\binom{m(n-n_1)+k-k_1-1}{k-k_1-1}\right)\\ -\frac{ml}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-2}{k-1}$$

for (17b). In the triple sum, we replace n_1 by $n_1+k_1+l_1+1$. Forgetting the sum over l_1, we see that it is then in the form of the left-hand side of (8) with n replaced by $n-k-l$, $a = m+1$, $c = m$, $\alpha = ml_1$, $\alpha' = m(l-l_1)$, $b = d = 1$, $\beta = l_1+1$, $\beta' = l-l_1$, and k replaced by $k-1$. Hence, if we substitute the right-hand side, the expression simplifies to

$$m\sum_{l_1=0}^{l-1}\binom{n-1}{k+l-1}\binom{m(n-1)+k-2}{k-1}\\ -\frac{ml}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-2}{k-1}. \tag{19}$$

Also here, the sum over l_1 sums the same summand for each l_1, so that the result is that summand multiplied by l.

Finally we treat the sum (17c). Again, we want to extend the sum over n_1 to $n_1 \ge 0$. In order to do so, we would have to subtract the terms for $n_1 = 1$ and $n_1 = 0$. However, it is somewhat unclear which values we should give the summand for these choices of n_1. To obtain a partial answer, we rewrite

$$\frac{1}{n_1-1}\binom{n_1-1}{k_1-1}\binom{m(n_1-1)+k_1-1}{k_1}\\ =\frac{m}{m(n_1-1)+k_1}\binom{n_1-1}{n_1-k_1}\binom{m(n_1-1)+k_1}{k_1}. \tag{20}$$

(This rewriting is already in the spirit of the forth-coming application of Carlitz's identity (7). We alert the reader that, according to our convention (6), the rewriting $\binom{n_1-1}{k_1-1}=\binom{n_1-1}{n_1-k_1}$ is without problem as long as $n_1\ge 1$, which is the case in (17c). However, it becomes wrong if $1>n_1\ge k_1$, in which case $\binom{n_1-1}{k_1-1}=0$ while $\binom{n_1-1}{n_1-k_1}\ne 0$. In the following considerations, whenever we talk about cases where $1>n_1\ge k_1$, we shall talk about the *right-hand side* in (20).) If $n_1=0$, then, because of the presence of the binomial coefficient $\binom{n_1-1}{n_1-k_1}$, the summand is only non-zero if $k_1=0$, in which case it equals

$$-\frac{ml}{mn+k}\binom{n}{n-k-l}\binom{mn+k}{k}=-\frac{l}{k+l}\binom{n-1}{k+l-1}\binom{mn+k-1}{k}.$$

For $n_1=1$, the above expression vanishes certainly if $k_1>1$. If $k_1=1$, it is equal to m. But if $k_1=0$, it is still not clear which value to assign to it. Leaving this question open for the moment, the arguments so far show that the expression (17c) is equal to

$$\begin{aligned}-\sum_{k_1=0}^{k}\sum_{n_1=0}^{n-1}\Bigg(&\frac{m}{m(n_1-1)+k_1}\binom{n_1-1}{n_1-k_1}\binom{m(n_1-1)+k_1}{k_1}\\ &\cdot\frac{ml}{m(n-n_1)+k-k_1}\binom{n-n_1}{n-n_1-k+k_1-l}\binom{m(n-n_1)+k-k_1}{k-k_1}\Bigg)\\ -\frac{l}{k+l}\binom{n-1}{k+l-1}&\binom{mn+k-1}{k}+\frac{ml}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-2}{k-1}\\ &+\ (\text{summand for } n_1=1,\ k_1=0).\end{aligned}$$

We now replace n_1 by n_1+k_1 in the double sum. This leads to the expression

$$\begin{aligned}-\sum_{n_1,k_1\ge 0}\Bigg(&\frac{m}{m(n_1+k_1-1)+k_1}\binom{n_1+k_1-1}{n_1}\binom{m(n_1+k_1-1)+k_1}{k_1}\\ &\cdot\frac{ml}{m(n-n_1-k_1)+k-k_1}\binom{n-n_1-k_1}{n-n_1-k-l}\binom{m(n-n_1-k_1)+k-k_1}{k-k_1}\Bigg)\\ -\frac{l}{k+l}\binom{n-1}{k+l-1}&\binom{mn+k-1}{k}+\frac{ml}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-2}{k-1}\\ &+\ (\text{summand for } n_1=1,\ k_1=0).\end{aligned}$$

The double sum is now exactly equal to the negative of the left-hand side of (7) with n replaced by $n-k-l$, $a=m+1$, $c=m$, $\alpha=-m$, $\alpha'=ml$, $b=d=1$, $\beta=-1$, $\beta'=l$. From there, we can also determine the missing value of the summand for $n_1=1$ and $k_1=0$. Namely, we have

$$A_{0,1}(\alpha,\beta)=\beta=-1.$$

Thus, if we substitute the right-hand side of (7), we obtain

$$\frac{l-1}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-1}{k}-\frac{l}{k+l}\binom{n-1}{k+l-1}\binom{mn+k-1}{k}$$
$$+\frac{ml}{k+l-1}\binom{n-2}{k+l-2}\binom{m(n-1)+k-2}{k-1}$$
$$+\frac{l}{k+l}\binom{n-2}{k+l-1}\binom{m(n-1)+k-1}{k} \tag{21}$$

for (17c).

Adding the expressions (18), (19), (21) and (17d), we obtain that the sum in (17) is equal to

$$l\binom{n}{k+l}\binom{m(n-1)+k-1}{k}+ml\binom{n-1}{k+l-1}\binom{m(n-1)+k-2}{k-1},$$

which is indeed equal to $l\cdot f_{k,l}(D_n,m)$ if $l\geq 1$. This proves (15).

Next we compute the sum on the left-hand side of (16),

$$\sum_{k_1=0}^{k}\binom{n}{k}\binom{m(n-1)+k_1-1}{k_1}+m\sum_{k_1=0}^{k}\binom{n-1}{k-1}\binom{m(n-1)+k_1-2}{k_1-1}$$
$$-\frac{1}{n-1}\binom{n-1}{k-1}\binom{m(n-1)+k-1}{k}$$
$$=\binom{n}{k}\binom{m(n-1)+k}{k}+m\binom{n-1}{k-1}\binom{m(n-1)+k-1}{k-1}$$
$$-\frac{1}{n-1}\binom{n-1}{k-1}\binom{m(n-1)+k-1}{k}$$
$$=\binom{n}{k}\binom{m(n-1)+k}{k}+\binom{n-2}{k-2}\binom{m(n-1)+k-1}{k},$$

the simplification of summations being due to the Chu–Vandermonde summation. This completes the proof. □

7 The *F*-triangle in the Exceptional Cases

It is a routine matter to use Proposition F in Section 2 (and a computer algebra package) to find the F-triangles for the exceptional root systems. We list our findings below.

The F-triangle for $I_2(a)$:

$$F^m_{I_2(a)}(x,y)=\frac{m(ma+a-2)}{2}x^2+2mxy+amx+y^2+2y+1. \tag{22}$$

The F-triangle for H_3:

$$F_{H_3}^m(x,y) = \frac{m(5m+2)(5m+4)}{3}x^3 + m(5m+2)x^2y + 5m(5m+2)x^2 + 3mxy^2 + 10mxy + 15mx + y^3 + 3y^2 + 3y + 1. \quad (23)$$

The F-triangle for H_4:

$$F_{H_4}^m(x,y) = \frac{m(3m+1)(5m+3)(15m+14)}{4}x^4 + m(3m+1)(5m+3)x^3y + 15m(3m+1)(5m+3)x^3 + \frac{1}{2}m(17m+5)x^2y^2 + m(45m+14)x^2y + \frac{1}{2}m(465m+149)x^2 + 4mxy^3 + 17mxy^2 + 31mxy + 60mx + y^4 + 4y^3 + 6y^2 + 4y + 1. \quad (24)$$

The F-triangle for F_4:

$$F_{F_4}^m(x,y) = \frac{m(2m+1)(3m+1)(6m+5)}{2}x^4 + 2m(2m+1)(3m+1)x^3y + 12m(2m+1)(3m+1)x^3 + 2m(4m+1)x^2y^2 + 2m(18m+5)x^2y + m(78m+23)x^2 + 4mxy^3 + 16mxy^2 + 26mxy + 24mx + y^4 + 4y^3 + 6y^2 + 4y + 1. \quad (25)$$

The F-triangle for E_6:

$$\begin{aligned}F_{E_6}^m(x,y) = {} & \frac{1}{30}m(2m+1)(3m+1)(4m+1)(6m+5)(12m+7)x^6 \\ & + \frac{1}{5}m(2m+1)(3m+1)(4m+1)(12m+7)x^5y \\ & + \frac{6}{5}m(2m+1)(3m+1)(4m+1)(12m+7)x^5 \\ & + \frac{1}{2}m(3m+1)(4m+1)(8m+3)x^4y^2 + 2m(3m+1)(4m+1)(12m+5)x^4y \\ & + 2m(3m+1)(4m+1)(30m+13)x^4 + \frac{5}{3}m(4m+1)(5m+1)x^3y^3 \\ & + m(4m+1)(48m+11)x^3y^2 + m(4m+1)(120m+31)x^3y \\ & + 9m(4m+1)(18m+5)x^3 + \frac{5}{2}m(7m+1)x^2y^4 + 5m(20m+3)x^2y^3 \\ & + m(242m+39)x^2y^2 + 3m(108m+19)x^2y + 12m(21m+4)x^2 + 6mxy^5 \\ & + 35mxy^4 + 85mxy^3 + 111mxy^2 + 84mxy + 36mx \\ & + y^6 + 6y^5 + 15y^4 + 20y^3 + 15y^2 + 6y + 1. \quad (26)\end{aligned}$$

The F-triangle for E_7:

$$\begin{aligned}
F^m_{E_7}(x,y) = {} & \frac{1}{280}m(3m+1)(3m+2)(9m+2)(9m+4)(9m+5)(9m+8)x^7 \\
&+\frac{9}{40}m(3m+1)(3m+2)(9m+2)(9m+4)(9m+5)x^6 \\
&+\frac{1}{40}m(3m+1)(3m+2)(9m+2)(9m+4)(9m+5)x^6y \\
&+\frac{3}{40}m(3m+1)(7m+3)(9m+2)(9m+4)x^5y^2 \\
&+\frac{3}{20}m(3m+1)(9m+2)(9m+4)(27m+13)x^5y \\
&+\frac{3}{40}m(3m+1)(9m+2)(9m+4)(207m+103)x^5 \\
&+\frac{1}{8}m(3m+1)(9m+2)(27m+7)x^4y^3 \\
&+\frac{3}{8}m(3m+1)(9m+2)(63m+19)x^4y^2 \\
&+\frac{3}{8}m(3m+1)(9m+2)(207m+71)x^4y \\
&+\frac{21}{8}m(3m+1)(9m+2)(63m+23)x^4+m(6m+1)(9m+2)x^3y^4 \\
&+\frac{3}{2}m(9m+2)(27m+5)x^3y^3+\frac{3}{2}m(9m+2)(81m+17)x^3y^2 \\
&+\frac{21}{2}m(9m+2)(21m+5)x^3y+\frac{21}{2}m(9m+2)(27m+7)x^3 \\
&+3m(8m+1)x^2y^5+3m(54m+7)x^2y^4+\frac{3}{2}m(315m+43)x^2y^3 \\
&+\frac{21}{2}m(75m+11)x^2y^2+\frac{21}{2}m(81m+13)x^2y+\frac{21}{2}m(63m+11)x^2 \\
&+7mxy^6+48mxy^5+141mxy^4+231mxy^3+231mxy^2+147mxy \\
&+63mx+y^7+7y^6+21y^5+35y^4+35y^3+21y^2+7y+1. \qquad (27)
\end{aligned}$$

The F-triangle for E_8:

$$\begin{aligned}
F^m_{E_8}(x,y) = {} & \frac{m(3m+1)(5m+1)(5m+2)(5m+3)(15m+8)(15m+11)(15m+14)x^8}{1344} \\
&+\frac{5}{56}m(3m+1)(5m+1)(5m+2)(5m+3)(15m+8)(15m+11)x^7 \\
&+\frac{1}{168}m(3m+1)(5m+1)(5m+2)(5m+3)(15m+8)(15m+11)yx^7 \\
&+\frac{1}{48}m(3m+1)(5m+1)(5m+2)(15m+7)(15m+8)y^2x^6 \\
&+\frac{5}{48}m(3m+1)(5m+1)(5m+2)(15m+8)(195m+107)x^6 \\
&+\frac{5}{24}m(3m+1)(5m+1)(5m+2)(15m+8)^2yx^6 \\
&+\frac{1}{3}m(3m+1)(5m+1)(5m+2)(10m+3)y^3x^5
\end{aligned}$$

$$
\begin{aligned}
&+\frac{5}{8}m(3m+1)(5m+1)(5m+2)(45m+16)y^2x^5\\
&+15m(3m+1)(5m+1)(5m+2)(30m+13)x^5\\
&+\frac{25}{8}m(3m+1)(5m+1)(5m+2)(39m+16)yx^5\\
&+\frac{1}{6}m(5m+1)(10m+3)(19m+4)y^4x^4+m(5m+1)(10m+3)(25m+6)y^3x^4\\
&+\frac{1}{4}m(5m+1)\left(3675m^2+2125m+308\right)y^2x^4\\
&+\frac{1}{2}m(5m+1)\left(10350m^2+6675m+1084\right)x^4\\
&+m(5m+1)\left(2250m^2+1395m+218\right)yx^4+\frac{7}{3}m(5m+1)(7m+1)y^5x^3\\
&+\frac{1}{3}m(5m+1)(380m+59)y^4x^3+\frac{1}{3}m(5m+1)(1315m+226)y^3x^3\\
&+m(5m+1)(915m+178)y^2x^3+45m(5m+1)(45m+11)x^3\\
&+m(5m+1)(1380m+307)yx^3+\frac{7}{2}m(9m+1)y^6x^2+7m(35m+4)y^5x^2\\
&+\frac{1}{2}m(1675m+199)y^4x^2+4m(415m+52)y^3x^2+\frac{1}{2}m(4295m+579)y^2x^2\\
&+\frac{35}{2}m(105m+17)x^2+75m(27m+4)yx^2+8my^7x+63my^6x+217my^5x\\
&+428my^4x+532my^3x+435my^2x+120mx+245myx\\
&+y^8+8y^7+28y^6+56y^5+70y^4+56y^3+28y^2+8y+1. \qquad (28)
\end{aligned}
$$

8 The $F = M$ Conjecture

In order to state the "$F = M$ Conjecture" for generalised cluster complexes, we need to first introduce Armstrong's [Arm06$^+$] m-divisible non-crossing partition posets.

Given a root system Φ and an element $\alpha \in \Phi$, let t_α denote the corresponding reflection in the central hyperplane perpendicular to α. Let $W = W(\Phi)$ be the group generated by these reflections. By definition, any element w of W can be represented as a product $w = t_1 t_2 \cdots t_\ell$, where the t_i's are reflections. We call the minimal number of reflections which is needed for such a product representation the *absolute length* of w, and we denote it by $\ell_T(w)$. We then define the *absolute order* on W, denoted by $\le_T$, by

$$u \le_T w \quad \text{if and only if} \quad \ell_T(w) = \ell_T(u) + \ell_T(u^{-1}w).$$

It can be shown that this is equivalent to the statement that any shortest product representation of u by reflections occurs as an initial segment in some shortest product representation of w by reflections.

We can now define the *non-crossing partition lattice* $NC(\Phi)$. Let c be a *Coxeter element* in W, that is, the product of all reflections corresponding to the simple roots. Then $NC(\Phi)$ is defined to be the restriction of the partial order $\le_T$ to the set of all elements which are less than or equal to c in absolute order. This definition makes sense because, regardless of the chosen Coxeter element c, the resulting poset is always the same up to isomorphism. It can be shown that $NC(\Phi)$ is indeed a lattice. (See [BW06$^+$] for a uniform proof.) The term "non-crossing partition lattice" is used because $NC(A_n)$ is isomorphic to the lattice of non-crossing partitions originally introduced by Kreweras [Kre72], and because also $NC(B_n)$ and $NC(D_n)$ can be realized as lattices of non-crossing partitions (see [AR04, Rei97]).

The poset of *m-divisible non-crossing partitions* has as a groundset the following subset of $(NC(\Phi))^{m+1}$,

$$NC^m(\Phi) = \{(w_0; w_1, \dots, w_m) : w_0 w_1 \cdots w_m = c \text{ and } \ell_T(w_0) + \ell_T(w_1) + \ldots + \ell_T(w_m) = \ell_T(c)\}. \tag{29}$$

The order relation is defined by

$$(u_0; u_1, \dots, u_m) \le (w_0; w_1, \dots, w_m) \quad \text{if and only if} \quad u_i \ge_T w_i,\ 1 \le i \le m.$$

We emphasize that, according to this definition, u_0 and w_0 need not be related in any way. The poset $NC^m(\Phi)$ is graded by the rank function

$$\operatorname{rk}\big((w_0; w_1, \dots, w_m)\big) = \ell_T(w_0).$$

Thus, there is a unique maximal element, namely $(c; \varepsilon, \dots, \varepsilon)$, where ε stands for the identity element in W, but, if $m > 1$, there are many different minimal elements. In particular, there is no global minimum in $NC^m(\Phi)$ if $m > 1$ and, hence, $NC^m(\Phi)$ is not a lattice for $m > 1$. (It is, however, a graded join-semilattice, see [Arm06$^+$, Sec. 2].) We remark that for $NC^m(A_n)$ and $NC^m(B_n)$ combinatorial realisations are available as subposets of non-crossing partitions in which each block has a size which is divisible by m. The corresponding translations are due to Armstrong [Arm06$^+$, Sec. 3]. In type A_n, the resulting poset had been earlier studied by Edelman [Ede81]. The analogous combinatorial realisation of $NC^m(D_n)$, generalising the one of Athanasiadis and Reiner [AR04] for $m = 1$, has not yet been worked out.

Next, we define the "M-triangle" of $NC^m(\Phi)$ as

$$M_\Phi^m(x, y) = \sum_{u, w \in NC^m(\Phi)} \mu(u, w)\, x^{\operatorname{rk} u} y^{\operatorname{rk} w},$$

where $\mu(u, w)$ is the Möbius function in $NC^m(\Phi)$.

The generalised version of Chapoton's (ex-)conjecture [Cha04, Conjecture 1], due to Armstrong [Arm06$^+$, Sec. 4], is the following.

Conjecture FM. *For any finite root system Φ of rank n, we have*

$$F_\Phi^m(x, y) = y^n\, M_\Phi^m\left(\frac{1+y}{y-x}, \frac{y-x}{y}\right).$$

Equivalently,

$$(1-xy)^n F_\Phi^m\left(\frac{x(1+y)}{1-xy},\frac{xy}{1-xy}\right)=\sum_{u,w\in(NC^m(\Phi))^*}\mu^*(u,w)\,(-x)^{\operatorname{rk}^* w}(-y)^{\operatorname{rk}^* u}, \tag{30}$$

where $(NC^m(\Phi))^*$ *denotes the poset* dual *to* $NC^m(\Phi)$ (*i.e., the poset which arises from* $NC^m(\Phi)$ *by reversing all order relations*), *where* μ^* *denotes the Möbius function in* $(NC^m(\Phi))^*$, *and where* rk^* *denotes the rank function in* $(NC^m(\Phi))^*$.

Since the Möbius function is multiplicative (see e.g. [Sta98, Prop. 3.8.2]), the multiplicativity property (2) for the F-triangle holds also for the M-triangle. Therefore, it is enough to prove the conjecture for the irreducible root systems. In Sections 9–17, we provide proofs for the root systems of type A_n, B_n, $I_2(a)$, H_3, H_4, F_4, and E_6, and a partial proof for the root system of type D_n.

9 Proof of the $F = M$ Conjecture for A_n

In this section we prove Conjecture FM for the type A_n. In the spirit of this paper, we follow a computational approach. We first simplify the left-hand side of (30) by a double application of the Chu–Vandermonde summation. Subsequently, we compute the right-hand side of (30) by relying on a result on rank selected chain enumeration in the m-divisible non-crossing partition lattice in type A_n due to Edelman [Ede81].

The link between chain enumeration and the Möbius function is the following. (The reader should consult [Sta98, Sec. 3.11] for more information on this topic.) Given a poset P and two elements u and w, $u \le w$, in the poset, the *zeta polynomial* of the interval $[u,w]$, denoted by $Z(u,w;z)$, is the number of (multi)chains from u to w of length z. (It can be shown that this is indeed a polynomial in z.) Then the Möbius function of u and w is equal to $\mu(u,w) = Z(u,w;-1)$.

Proposition A. *In type* A_n, *the left-hand side of (30) is equal to*

$$\sum_{r,s\ge 0} x^s y^r \frac{1}{s+1}\binom{n}{s}\binom{m(n+1)}{r}\binom{m(n+1)+s-r-1}{s-r}. \tag{31}$$

Proof. By definition of $F_{A_n}^m(x,y)$, and by Theorem FA in Section 4, the left-hand side of (30) in type A_n is equal to

$$\sum_{k,l,r,s\ge 0}\frac{l+1}{k+l+1}\binom{n}{k+l}\binom{m(n+1)+k-1}{k}\binom{k}{r}\binom{n-k-l}{s}(-1)^s x^{k+l+s}y^{l+r+s}.$$

Fixing $L = s + l$, we rewrite this as

$$\sum_{k,L,r\ge 0}\frac{n!\,(m(n+1)+k-1)!}{(m(n+1)-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L+1)!}x^{k+L}y^{r+L} \cdot\sum_{s=0}^{L}(L-s+1)(-1)^s\binom{k+L+1}{s}.$$

We compute the sum over s by the Chu–Vandermonde summation. Thus, we arrive at

$$\sum_{k,L,r\ge 0}\frac{n!\,(m(n+1)+k-1)!}{(m(n+1)-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L+1)!}x^{k+L}y^{r+L} \cdot(-1)^L\binom{k+L-1}{L}.$$

We now write $K=k+L$ and $R=r+L$. Subsequently, the sum over L can be computed using the Chu–Vandermonde summation. The result is

$$\sum_{K,R\ge 0}\frac{n!\,(m(n+1)+K-R-1)!\,(m(n+1))!}{(m(n+1)-1)!\,(m(n+1)-R)!\,R!\,(K-R)!\,(n-K)!\,(K+1)!}x^K y^R.$$

Aside from a parameter replacement, this is exactly the expression (31). ☐

For the computation of the right-hand side of (30) we require the following theorem due to Edelman [Ede81].

Theorem NA. *The number of chains in $(NC^m(A_n))^*$ with successive rank jumps $s_1,s_2,\dots,s_\ell$, $s_1+s_2+\dots+s_\ell=n$, is*

$$\binom{m(n+1)}{s_1}\cdots\binom{m(n+1)}{s_{\ell-1}}\frac{1}{n+1}\binom{n+1}{s_\ell}. \tag{32}$$

Proof of Conjecture FM in type A_n. We now compute the right-hand side of (30), that is,

$$\sum_{u,w\in(NC^m(A_n))^*}\mu^*(u,w)(-x)^{\operatorname{rk}^* w}(-y)^{\operatorname{rk}^* u}.$$

In order to compute the coefficient of $x^s y^r$ in this expression,

$$(-1)^{r+s}\sum_{\substack{u,w\in(NC^m(A_n))^*\\ \text{with }\operatorname{rk}^* u=r\text{ and }\operatorname{rk}^* w=s}}\mu^*(u,w),$$

we compute the sum of all corresponding zeta polynomials (in the variable z), multiplied by $(-1)^{r+s}$,

$$(-1)^{r+s}\sum_{\substack{u,w\in(NC^m(A_n))^*\\ \text{with }\operatorname{rk}^* u=r\text{ and }\operatorname{rk}^* w=s}}Z(u,w;z),$$

and then put $z=-1$.

For computing this sum of zeta polynomials, we must set $\ell = z+2$, $s_1 = r$, $n - s_\ell = s$, $s_2 + s_3 + \ldots + s_{\ell-1} = s - r$ in (32), and then sum the resulting expression over all possible $s_2, s_3, \ldots, s_{\ell-1}$. By using the Chu–Vandermonde summation, one obtains

$$\frac{1}{n+1}\binom{m(n+1)}{r}\binom{zm(n+1)}{s-r}\binom{n+1}{n-s}.$$

If we put $z = -1$ in this expression and multiply it by $(-1)^{r+s}$, then we obtain exactly the coefficient of $x^s y^r$ in (31). □

10 Proof of the $F = M$ Conjecture for B_n

In this section we prove Conjecture FM for the type B_n, by following the same approach as the one for type A_n in the previous section.

Proposition B. *In type B_n, the left-hand side of (30) is equal to*

$$\sum_{r,s\ge 0} x^s y^r \binom{n}{s}\binom{mn}{r}\binom{mn+s-r-1}{s-r}. \tag{33}$$

Proof. By definition of $F^m_{B_n}(x,y)$, and by Theorem FB in Section 5, the left-hand side of (30) in type B_n is equal to

$$\sum_{k,l,r,s\ge 0}\binom{n}{k+l}\binom{mn+k-1}{k}\binom{k}{r}\binom{n-k-l}{s}(-1)^s x^{k+l+s} y^{l+r+s}.$$

Fixing $L = s + l$, we rewrite this as

$$\sum_{k,L,r\ge 0}\frac{n!\,(mn+k-1)!}{(mn-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L)!}x^{k+L}y^{L+r}\sum_{s=0}^{L}(-1)^s\binom{k+L}{s}.$$

We compute the sum over s by the Chu–Vandermonde summation. Thus, we arrive at

$$\sum_{k,L,r\ge 0}\frac{n!\,(mn+k-1)!}{(mn-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L)!}x^{k+L}y^{L+r}(-1)^L\binom{k+L-1}{L}.$$

We now write $K = k + L$ and $R = r + L$. Subsequently, the sum over L can be computed using the Chu–Vandermonde summation. The result is

$$\sum_{K,R\ge 0}\frac{n!\,(mn+K-R-1)!\,(mn)!}{(mn-1)!\,(mn-R)!\,R!\,(K-R)!\,(n-K)!\,K!}x^K y^R.$$

Aside from a parameter replacement, this is exactly the expression (33). □

For the computation of the right-hand side of (30) we require the following theorem due to Armstrong [Arm06⁺, Sec. 3].

Theorem NB. *The number of chains in* $(NC^m(B_n))^*$ *with successive rank jumps* $s_1, s_2, \dots, s_\ell$, $s_1 + s_2 + \dots + s_\ell = n$, *is*

$$\binom{mn}{s_1} \cdots \binom{mn}{s_{\ell-1}} \binom{n}{s_\ell}. \tag{34}$$

Proof of Conjecture FM in type B_n. We now compute the right-hand side of (30), that is,

$$\sum_{u,w \in (NC^m(B_n))^*} \mu^*(u,w)(-x)^{\operatorname{rk}^* w}(-y)^{\operatorname{rk}^* u}.$$

In order to compute the coefficient of $x^s y^r$ in this expression,

$$(-1)^{r+s} \sum_{\substack{u,w \in (NC^m(B_n))^* \\ \text{with } \operatorname{rk}^* u = r \text{ and } \operatorname{rk}^* w = s}} \mu^*(u,w),$$

we compute the sum of all corresponding zeta polynomials (in the variable z), multiplied by $(-1)^{r+s}$,

$$(-1)^{r+s} \sum_{\substack{u,w \in (NC^m(B_n))^* \\ \text{with } \operatorname{rk}^* u = r \text{ and } \operatorname{rk}^* w = s}} Z(u,w;z),$$

and then put $z = -1$.

For computing this sum of zeta polynomials, we must set $\ell = z + 2$, $s_1 = r$, $n - s_\ell = s$, $s_2 + s_3 + \dots + s_{\ell-1} = s - r$ in (34), and then sum the resulting expression over all possible $s_2, s_3, \dots, s_{\ell-1}$. By using the Chu–Vandermonde summation, one obtains

$$\binom{mn}{r} \binom{zmn}{s-r} \binom{n}{n-s}.$$

If we put $z = -1$ in this expression and multiply it by $(-1)^{r+s}$, then we obtain exactly the coefficient of $x^s y^r$ in (33). ◻

11 Towards a Proof of the $F = M$ Conjecture for D_n

This section exhibits how far the approach of the previous two sections of proving Conjecture FM in types A_n and B_n can take us in type D_n. The simplification of the left-hand side of (30) along the lines of the proofs of Propositions A and B goes through smoothly. The problem which we face in type D_n is that, up to this date, the rank selected chain enumeration result for

$NC^m(D_n)$ has not been found yet. Thus, we do not have the means to compute the M-triangle for $NC^m(D_n)$. The only exception is for $m = 1$. Namely, for the (ordinary) non-crossing partition lattice $NC(D_n) = NC^1(D_n)$, Athanasiadis and Reiner [AR04] have done the rank selected chain enumeration as we need it in our application. Hence, we are able to prove Conjecture FM in type D_n if $m = 1$.

Proposition D. *In type D_n, the left-hand side of (30) is equal to*

$$\sum_{r,s\ge 0} x^s y^r \Bigg(2\binom{n-1}{s-1}\binom{m(n-1)}{r}\binom{m(n-1)+s-r-1}{s-r} + \binom{n-2}{s}\binom{m(n-1)}{r}\binom{m(n-1)+s-r-1}{s-r} + m\binom{n-1}{s-1}\binom{m(n-1)-1}{r-2}\binom{m(n-1)+s-r-1}{s-r} - m\binom{n-1}{s-1}\binom{m(n-1)}{r}\binom{m(n-1)+s-r-2}{s-r-2}\Bigg). \tag{35}$$

Proof. By definition of $F^m_{D_n}(x,y)$, and by Theorem FD in Section 6, the left-hand side of (30) in type D_n is equal to

$$\sum_{k,l,r,s\ge 0} \binom{n}{k+l}\binom{m(n-1)+k-1}{k}\binom{k}{r}\binom{n-k-l}{s}(-1)^s x^{k+l+s} y^{l+r+s} \tag{36a}$$

$$+m\sum_{k,l,r,s\ge 0} \binom{n-1}{k+l-1}\binom{m(n-1)+k-2}{k-1}\binom{k}{r}\binom{n-k-l}{s}(-1)^s x^{k+l+s} y^{l+r+s} \tag{36b}$$

$$-\frac{1}{n-1}\sum_{k,r,s\ge 0} \binom{n-1}{k-1}\binom{m(n-1)+k-1}{k}\binom{k}{r}\binom{n-k}{s}(-1)^s x^{k+s} y^{r+s}. \tag{36c}$$

We treat the three sums in (36) separately. We begin with the sum (36a). Fixing $L = s + l$, we rewrite it as

$$\sum_{k,L,r\ge 0} \frac{n!\,(m(n-1)+k-1)!}{(m(n-1)-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L)!} x^{k+L} y^{L+r} \cdot \sum_{s=0}^{L} (-1)^s \binom{k+L}{s}.$$

We compute the sum over s by the Chu–Vandermonde summation. Thus, we arrive at

$$\sum_{k,L,r\ge 0} \frac{n!\,(m(n-1)+k-1)!}{(m(n-1)-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L)!} x^{k+L} y^{L+r} \cdot (-1)^L \binom{k+L-1}{L}.$$

We now write $K = k + L$ and $R = r + L$. Subsequently, the sum over L can be computed using the Chu–Vandermonde summation. The result is

$$\sum_{K,R\ge 0} \frac{n!\,(m(n-1)+K-R-1)!\,(m(n-1))!}{(m(n-1)-1)!\,(m(n-1)-R)!\,R!\,(K-R)!\,(n-K)!\,K!} x^K y^R. \tag{37}$$

Next we consider the sum (36b). Again fixing $L = s + l$, we rewrite it as

$$\sum_{k,L,r\ge 0} \frac{mk\,(n-1)!\,(m(n-1)+k-2)!}{(m(n-1)-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L-1)!} x^{k+L} y^{L+r} \cdot \sum_{s=0}^{L} (-1)^s \binom{k+L-1}{s}.$$

We compute the sum over s by the Chu–Vandermonde summation. Thus, we arrive at

$$\sum_{k,L,r\ge 0} \frac{mk\,(n-1)!\,(m(n-1)+k-2)!}{(m(n-1)-1)!\,r!\,(k-r)!\,(n-k-L)!\,(k+L-1)!} x^{k+L} y^{L+r} \cdot (-1)^L \binom{k+L-2}{L}.$$

We now write $K = k + L$ and $R = r + L$. Because of the presence of the factor k in the numerator, this makes a factor of $K - L$ appear. We split the sum into two parts accordingly, and then, in both parts, the sum over L can be computed using the Chu–Vandermonde summation. The result is

$$\sum_{K,R\ge 0} \frac{mK\,(n-1)!\,(m(n-1)+K-R-2)!\,(m(n-1))!}{(m(n-1)-1)!\,(m(n-1)-R)!\,R!\,(K-R)!\,(n-K)!\,(K-1)!} x^K y^R - \sum_{K,R\ge 0} \frac{m(K-2)\,(n-1)!}{(m(n-1)-1)!\,(m(n-1)-R+1)!} \cdot \frac{(m(n-1)+K-R-2)!\,(m(n-1))!}{(R-1)!\,(K-R)!\,(n-K)!\,(K-1)!} x^K y^R. \tag{38}$$

Finally, we turn to the sum (36c). We write $K = k + s$ and $R = r + s$. Subsequently, the sum over s can be computed using the Chu–Vandermonde summation. The result is

$$-\sum_{K,R\ge 0} \frac{1}{n-1} \frac{(n-1)!\,(m(n-1)+K-R-1)!\,(m(n-1))!}{(m(n-1)-1)!\,(m(n-1)-R)!\,R!\,(K-R)!\,(n-K)!\,(K-1)!} x^K y^R. \tag{39}$$

To complete the proof, we add the expressions (37), (38) and (39). Doing the parameter replacements $K \to s$, $R \to r$ and minor rewriting, this leads to the expression

$$\sum_{r,s\ge 0} x^s y^r \Bigg(\binom{n}{s}\binom{m(n-1)}{r}\binom{m(n-1)+s-r-1}{s-r} \\ + \frac{ms}{s-r}\binom{n-1}{s-1}\binom{m(n-1)}{r}\binom{m(n-1)+s-r-2}{s-r-1} \\ + \frac{m(s-2)}{s-r}\binom{n-1}{s-1}\binom{m(n-1)}{r-1}\binom{m(n-1)+s-r-2}{s-r-1} \\ - \frac{1}{n-1}\binom{n-1}{s-1}\binom{m(n-1)}{r}\binom{m(n-1)+s-r-1}{s-r}\Bigg).$$

It is now a routine verification to show that an alternative way to write this is (35).

For the computation of the right-hand side of (30) in the case that $m = 1$, we require the following result due to Athanasiadis and Reiner [AR04, Theorem 1.2(ii)].

Theorem ND. *The number of chains in $NC(D_n)$ with successive rank jumps $s_1, s_2, \dots, s_\ell$, $s_1 + s_2 + \dots + s_\ell = n$, is given by*

$$2\binom{n-1}{s_1}\cdots\binom{n-1}{s_\ell} + \sum_{i=1}^{\ell}\binom{n-1}{s_1}\cdots\binom{n-2}{s_i-2}\cdots\binom{n-1}{s_\ell}. \qquad (40)$$

Proof of Conjecture FM for $m = 1$ in type D_n. If $m = 1$, then $NC^m(D_n)$ reduces to the ordinary non-crossing partition lattice $NC(D_n)$, which is self-dual, that is, $(NC(D_n))^* = NC(D_n)$. Hence, the right-hand side of (30) with $m = 1$ in type D_n is equal to

$$\sum_{u,w\in NC(D_n)} \mu(u,w)(-x)^{\operatorname{rk} w}(-y)^{\operatorname{rk} u}.$$

In order to compute the coefficient of $x^s y^r$ in this expression,

$$(-1)^{r+s} \sum_{\substack{u,w\in NC(D_n)\\ \text{with } \operatorname{rk} u = r \text{ and } \operatorname{rk} w = s}} \mu(u,w),$$

we compute the sum of all corresponding zeta polynomials (in the variable z), multiplied by $(-1)^{r+s}$,

$$(-1)^{r+s} \sum_{\substack{u,w\in NC(D_n)\\ \text{with } \operatorname{rk} u = r \text{ and } \operatorname{rk} w = s}} Z(u,w;z),$$

and then put $z = -1$.

For computing this sum of zeta polynomials, we must set $\ell = z+2$, $s_1 = r$, $n - s_\ell = s$, $s_2 + s_3 + \ldots + s_{\ell-1} = s - r$ in (40), and then sum the resulting expression over all possible $s_2, s_3, \ldots, s_{\ell-1}$. By using the Chu–Vandermonde summation again, one obtains

$$2\binom{n-1}{r}\binom{z(n-1)}{s-r}\binom{n-1}{n-s} + \binom{n-2}{r-2}\binom{z(n-1)}{s-r}\binom{n-1}{n-s}$$
$$+\binom{n-1}{r}\binom{z(n-1)}{s-r}\binom{n-2}{n-s-2} + z\binom{n-1}{r}\binom{z(n-1)-1}{s-r-2}\binom{n-1}{n-s}.$$

If we put $z = -1$ in this expression and multiply it by $(-1)^{r+s}$, then we obtain exactly the coefficient of $x^s y^r$ in (35) with $m = 1$. □

12 How to Prove the $F = M$ Conjecture in the Exceptional Cases

While, at first sight, for a given exceptional root system Φ, it seems that computing the M-triangle (respectively the right-hand side of (30)) for *arbitrary* m is an infinite problem because we have to compute Möbius functions for $NC^m(\Phi)$ (respectively for $(NC^m(\Phi))^*$) for $m = 1, 2, \ldots$, this is not really true. We should recall from (29) that an element of $NC^m(\Phi)$ has the form

$$(w_0; w_1, \ldots, w_m), \text{ with } w_0 w_1 \cdots w_m = c \text{ and}$$
$$\ell_T(w_0) + \ell_T(w_1) + \ldots + \ell_T(w_m) = \ell_T(c) = n, \qquad (41)$$

where n is the rank of the root system Φ. Now, n can be at most 8 for an exceptional root system (with equality only for $\Phi = E_8$). This implies that only at most 8 of the w_i's can be different from the identity element ε in $W = W(\Phi)$. Hence, a typical interval in $(NC^m(\Phi))^*$ looks like $[u, w]$, where

$$u = (u_0; u_1, \ldots, u_m), \quad w = (w_0; w_1, \ldots, w_m),$$

$u_i = w_i = \varepsilon$ for all but at most 8 indices $i \geq 1$, and $u_i \leq w_i$ for these remaining indices. Let these latter indices be $i_1, i_2, \ldots, i_d$, with $d \leq 8$. Then, such an interval $[u, w]$ is isomorphic to the "compressed" interval $[u', w']$, where

$$u' = (u_0; u_{i_1}, \ldots, u_{i_d}), \quad w' = (w_0; w_{i_1}, \ldots, w_{i_d}).$$

Note that "compressed" means that all of $w_{i_1}, w_{i_2}, \ldots, w_{i_d}$ are different from ε.

So, what we have to do is to determine all different *compressed* intervals $[u', w']$. The contribution of a compressed interval $[u', w']$ to the right-hand side of (30) is then

$$\binom{m}{d} \cdot \mu^*(u', w')\, x^{\operatorname{rk}^* w'} y^{\operatorname{rk}^* u'}, \qquad (42)$$

because there are $\binom{m}{d}$ different ways to choose $\{i_1, i_2,\dots, i_d\}$ out of $\{1, 2,\dots,m\}$. To obtain the M-triangle we "just" have to collect all these contributions and sum them over all possible compressed intervals. Note that this is now a *finite* problem because the number of *compressed* intervals is finite.

Rather than running through all compressed intervals, a more efficient way to implement this is as follows. We rewrite the right-hand side of (30) as

$$\sum_{u,w\in(NC^m(\Phi))^*} \mu^*(u,w)\,(-x)^{\operatorname{rk}^* w}(-y)^{\operatorname{rk}^* u} = \sum_{w\in(NC^m(\Phi))^*} (-x)^{\operatorname{rk}^* w}\cdot\chi^*_{[\hat{0},w]}(-y), \tag{43}$$

where $\hat{0} = (c; \varepsilon, \dots, \varepsilon)$ is the minimum in $(NC^m(\Phi))^*$, and where

$$\chi^*_{[\hat{0},w]}(y) = \sum_{u\in(NC^m(\Phi))^*} \mu^*(u,w)\,y^{\operatorname{rk}^* u}$$

is, essentially, the characteristic polynomial of the interval $[\hat{0}, w]$. (To be precise, it is the characteristic polynomial of the interval $[w, \hat{0}]$ in $NC^m(\Phi)$, see [Sta98, Sec. 3.10].) If $w = (w_0; w_1, \dots, w_m)$ with $w_{i_1}, w_{i_2}, \dots, w_{i_d}$ those among $w_1, w_2, \dots, w_m$ which are different from the identity element ε, then

$$[\hat{0}, w] \cong [\varepsilon, w_{i_1}] \times [\varepsilon, w_{i_2}] \times \dots \times [\varepsilon, w_{i_d}],$$

where each interval $[\varepsilon, w_{i_j}]$ is an interval in $NC(\Phi)$. Since the characteristic polynomial is multiplicative, this implies

$$\chi^*_{[\hat{0},w]}(y) = \chi^*_{[\varepsilon,w_{i_1}]}(y)\chi^*_{[\varepsilon,w_{i_2}]}(y)\cdots\chi^*_{[\varepsilon,w_{i_d}]}(y),$$

where $\chi^*_{[\varepsilon,w_{i_j}]}(y) = \sum_{v\in NC(\Phi)} \mu(v, w_{i_j})y^{\operatorname{rk} v}$, with μ the Möbius function and rk the rank function in $NC(\Phi)$.

According to a result by Bessis [Bes03, Lemma 1.4.3, Cor. 1.6.2], each element w_{i_j} is some parabolic Coxeter element (that is, a Coxeter element in some parabolic subgroup), and the interval $[\varepsilon, w_{i_j}]$ is isomorphic to some $NC(\Psi)$, where Ψ is the root system of this parabolic subgroup.

If we put all this together, then (43) becomes

$$\sum_{d=0}^{n}\sum_{(T_1,\dots,T_d)} (-x)^{\operatorname{rk} T_1+\dots+\operatorname{rk} T_d}\cdot N_\Phi(T_1, T_2, \dots, T_d) \\ \cdot\chi^*_{NC(T_1)}(-y)\chi^*_{NC(T_2)}(-y)\cdots\chi^*_{NC(T_d)}(-y)\binom{m}{d}, \tag{44}$$

where the inner sum is over all possible d-tuples $(T_1, T_2, \dots, T_d)$ of types (not necessarily irreducible types), and where $N_\Phi(T_1, T_2, \dots, T_d)$ is the number of "minimal" products $c_1c_2\cdots c_d$ less than or equal to the Coxeter element c in absolute order, "minimal" meaning that all the c_i's are different from ε and that $\ell_T(c_1) + \ell_T(c_2) + \dots + \ell_T(c_d) = \ell_T(c_1c_2\cdots c_d)$, such that the type of c_i

as a parabolic Coxeter element is T_i, $i = 1, 2, \dots, d$. The notation $NC(T)$ in (44) means $NC(\Psi)$, where Ψ is a root system of type T, and $\operatorname{rk} T$ denotes the rank of Ψ. We point out that the appearance of the binomial coefficient $\binom{m}{d}$ is explained by (42).

So, what we have to do to apply formula (44) to compute the right-hand side of (30) is, first, to determine all the "decomposition numbers" $N_\Phi(T_1, T_2, \dots, T_d)$. Since we shall refer to it later, we point out that these decomposition numbers have many relations between themselves. For example, the number $N_\Phi(T_1, T_2, \dots, T_d)$ is independent of the order of the types $T_1, T_2, \dots, T_d$, that is, we have

$$N_\Phi(T_{\sigma(1)}, T_{\sigma(2)}, \dots, T_{\sigma(d)}) = N_\Phi(T_1, T_2, \dots, T_d) \tag{45}$$

for any permutation σ of $\{1, 2, \dots, d\}$. This follows from the (proof of the) Shifting Lemma (see [Arm06$^+$, Sec. 1]). Furthermore, by the definition of these numbers, those of "lower rank" can be computed from those of "full rank." To be precise, we have

$$N_\Phi(T_1, T_2, \dots, T_d) = \sum_T N_\Phi(T_1, T_2, \dots, T_d, T), \tag{46}$$

where the sum is over all types T of rank $n - \operatorname{rk} T_1 - \operatorname{rk} T_2 - \dots - \operatorname{rk} T_d$ (with n still denoting the rank of the fixed root system Φ).

Second, one needs a list of the characteristic polynomials $\chi^*_{NC(\Psi)}(y)$ for all *irreducible* root systems Ψ. (By the multiplicativity of the characteristic polynomial, this then gives also formulae for the characteristic polynomials of all the reducible types.) In fact, the numbers $N_\Psi(T_1, T_2, \dots, T_d)$ carry all the information which is necessary to do this recursively. Namely, by the definition of $NC(\Psi)$ and of the decomposition numbers $N_\Psi(T_1, T_2, \dots, T_d)$, we have

$$\chi^*_{NC(\Psi)}(y) = \sum_{T_1, T_2} N_\Psi(T_1, T_2)\, \mu_{NC(T_2)}\big(\hat{0}_{NC(T_2)}, \hat{1}_{NC(T_2)}\big)\, y^{\operatorname{rk} T_1}, \tag{47}$$

where $\mu_{NC(T_2)}(.,.)$ denotes the Möbius function in $NC(T_2)$, and where $\hat{0}_{NC(T_2)}$ and $\hat{1}_{NC(T_2)}$ are, respectively, the minimal and the maximal element in $NC(T_2)$. Indeed, inductively, the Möbius functions $\mu_{NC(T_2)}\big(\hat{0}_{NC(T_2)}, \hat{1}_{NC(T_2)}\big)$ are already known for all T_2 of lower rank than the rank of Ψ. Hence, the only unknown in (47) is $\mu_{NC(\Psi)}\big(\hat{0}_{NC(\Psi)}, \hat{1}_{NC(\Psi)}\big)$. However, the latter can be computed by setting $y = 1$ in (47) and using the fact that $\chi^*_{NC(\Psi)}(1) = 0$ for all root systems Ψ of rank at least 1. (This fact is equivalent to the statement that $\sum_{u \in NC(\Psi)} \mu_{NC(\Psi)}\big(u, \hat{1}_{NC(\Psi)}\big) = 0$, which is nothing but a part of the definition of the Möbius function. Alternatively, one may use the uniform formula for the zeta polynomial of the non-crossing partition lattices, in which one specializes the variable to -1. See [Cha04, Prop. 9]; the reader may be warned that a slightly different convention for the zeta polynomial is used there.)

We show in Sections 13–17 how to implement this procedure for the dihedral root system $I_2(a)$, for the hyperbolic root systems H_3 and H_4, and for F_4 and E_6. We list the values of the characteristic polynomials of the irreducible root systems that we need below.

$$\begin{aligned}
\chi^*_{A_1}(y) &= y-1,\\
\chi^*_{A_2}(y) &= y^2-3y+2,\\
\chi^*_{I_2(a)}(y) &= y^2-ay+a-1,\\
\chi^*_{A_3}(y) &= y^3-6y^2+10y-5,\\
\chi^*_{A_4}(y) &= y^4-10y^3+30y^2-35y+14,\\
\chi^*_{A_5}(y) &= y^5-15y^4+70y^3-140y^2+126y-42,\\
\chi^*_{B_3}(y) &= y^3-9y^2+18y-10,\\
\chi^*_{D_4}(y) &= y^4-12y^3+39y^2-48y+20,\\
\chi^*_{D_5}(y) &= y^5-20y^4+106y^3-230y^2+220y-77,\\
\chi^*_{H_3}(y) &= y^3-15y^2+35y-21,\\
\chi^*_{F_4}(y) &= y^4-24y^3+101y^2-144y+66,\\
\chi^*_{H_4}(y) &= y^4-60y^3+307y^2-480y+232,\\
\chi^*_{E_6}(y) &= y^6-36y^5+300y^4-1035y^3+1720y^2-1368y+418. \qquad (48)
\end{aligned}$$

13 Proof of the $F = M$ Conjecture for $I_2(a)$

By (22), we have

$$\begin{aligned}
(1-xy)^2 F^m_{I_2(a)}\left(\frac{x(1+y)}{1-xy},\frac{xy}{1-xy}\right) &= \frac{m(am-a+2)}{2}x^2y^2+am^2x^2y\\
&\quad+\frac{m(am+a-2)}{2}x^2+amxy+amx+1 \qquad (49)
\end{aligned}$$

for the left-hand side of (30).

We now compute the right-hand side of (30) following the proposed procedure in the previous section. We have $N_{I_2(a)}(I_2(a))=1$, $N_{I_2(a)}(A_1,A_1)=a$, $N_{I_2(a)}(A_1)=a$, $N_{I_2(a)}(\emptyset)=1$, all other numbers $N_{I_2(a)}(T_1,\dots,T_d)$ being zero. Thus, according to (44) and (48), the right-hand side of (30) is equal to

$$(-x)^2(y^2+ay+a-1)m+(-x)^2a(-y-1)^2\binom{m}{2}+(-x)a(-y-1)m+1,$$

which agrees with (49).

14 Proof of the $F = M$ Conjecture for H_3

By (23), we have

$$(1-xy)^3 F^m_{H_3}\left(\frac{x(1+y)}{1-xy}, \frac{xy}{1-xy}\right) = \frac{m(5m-4)(5m-2)}{3}x^3y^3 \\ +5m^2(5m-2)x^3y^2 + \frac{m(5m+2)(5m+4)}{3}x^3 \\ +5m^2(5m+2)x^3y + 5m(5m-2)x^2y^2 + 5m(5m+2)x^2 \\ +50m^2x^2y + 15mx + 15mxy + 1 \qquad (50)$$

for the left-hand side of (30).

We now compute the right-hand side of (30) following the proposed procedure in Section 12. The conclusions which we report here are based on *Maple* computations which we performed using Stembridge's `coxeter` package [Ste].

We have $N_{H_3}(H_3) = 1$, $N_{H_3}(A_1^2, A_1) = 5$, $N_{H_3}(A_2, A_1) = 5$, $N_{H_3}(I_2(5), A_1) = 5$, $N_{H_3}(A_1, A_1, A_1) = 50$, plus the assignments implied by (45) and (46), all other numbers $N_{H_3}(T_1, \dots, T_d)$ being zero. Thus, according to (44) and (48), the right-hand side of (30) is equal to

$$(-x)^3(-y^3 - 15y^2 - 35y - 21)m + 2\cdot(-x)^3 5(-y-1)^3\binom{m}{2} \\ +2\cdot(-x)^3 5(y^2+3y+2)(-y-1)\binom{m}{2} + 2\cdot(-x)^3 5(y^2+5y+4)(-y-1)\binom{m}{2} \\ +(-x)^3 50(-y-1)^3\binom{m}{3} + (-x)^2 5(-y-1)^2 m + (-x)^2 5(y^2+3y+2)m \\ +(-x)^2 5(y^2+5y+4)m + (-x)^2 50(-y-1)^2\binom{m}{2} + (-x)15(-y-1)m + 1,$$

which agrees with (50).

15 Proof of the $F = M$ Conjecture for H_4

By (24), we have

$$(1-xy)^4 F^m_{H_4}\left(\frac{x(1+y)}{1-xy}, \frac{xy}{1-xy}\right) = \frac{1}{4}m(3m-1)(5m-3)(15m-14)x^4y^4 \\ +15m^2(3m-1)(5m-3)x^4y^3 + \frac{1}{2}m^2\left(675m^2-61\right)x^4y^2 \\ +\frac{1}{4}m(3m+1)(5m+3)(15m+14)x^4 + 15m^2(3m+1)(5m+3)x^4y \\ +15m(3m-1)(5m-3)x^3y^3 + 15m^2(45m-14)x^3y^2 \\ +15m(3m+1)(5m+3)x^3 + 15m^2(45m+14)x^3y + \frac{1}{2}m(465m-149)x^2y^2 \\ +\frac{1}{2}m(465m+149)x^2 + 465m^2x^2y + 60mx + 60mxy + 1 \qquad (51)$$

for the left-hand side of (30).

We now compute the right-hand side of (30) following the proposed procedure in Section 12. The conclusions which we report here are based on *Maple* computations which we performed using Stembridge's `coxeter` package [Ste].

We have $N_{H_4}(H_4) = 1$, $N_{H_4}(A_1 * A_2, A_1) = 15$, $N_{H_4}(A_3, A_1) = 15$, $N_{H_4}(H_3, A_1) = 15$, $N_{H_4}(A_1 * I_2(5), A_1) = 15$, $N_{H_4}(A_1^2, A_1^2) = 30$, $N_{H_4}(A_1^2, A_2) = 30$, $N_{H_4}(A_1^2, I_2(5)) = 15$, $N_{H_4}(A_2, A_2) = 5$, $N_{H_4}(A_2, I_2(5)) = 15$, $N_{H_4}(I_2(5), I_2(5)) = 3$, $N_{H_4}(A_1^2, A_1, A_1) = 225$, $N_{H_4}(A_2, A_1, A_1) = 150$, $N_{H_4}(I_2(5), A_1, A_1) = 90$, $N_{H_4}(A_1, A_1, A_1, A_1) = 1350$, plus the assignments implied by (45) and (46), all other numbers $N_{H_4}(T_1, \dots, T_d)$ being zero. If one substitutes accordingly in (44), using the information from (48), then one obtains an expression which agrees with (51) after simplification. ◻

16 Proof of the $F = M$ Conjecture for F_4

By (25), we have

$$(1-xy)^4 F_{F_4}^m\left(\frac{x(1+y)}{1-xy}, \frac{xy}{1-xy}\right) = \frac{1}{2}m(2m-1)(3m-1)(6m-5)x^4y^4$$
$$+12m^2(2m-1)(3m-1)x^4y^3 + m^2\left(108m^2-7\right)x^4y^2$$
$$+\frac{1}{2}m(2m+1)(3m+1)(6m+5)x^4 + 12m^2(2m+1)(3m+1)x^4y$$
$$+12m(2m-1)(3m-1)x^3y^3 + 12m^2(18m-5)x^3y^2$$
$$+12m(2m+1)(3m+1)x^3 + 12m^2(18m+5)x^3y + m(78m-23)x^2y^2$$
$$+m(78m+23)x^2 + 156m^2x^2y + 24mx + 24mxy + 1 \tag{52}$$

for the left-hand side of (30).

We now compute the right-hand side of (30) following the proposed procedure in Section 12. The conclusions which we report here are based on *Maple* computations which we performed using Stembridge's `coxeter` package [Ste].

We have $N_{F_4}(F_4) = 1$, $N_{F_4}(A_1 * A_2, A_1) = 12$, $N_{F_4}(B_3, A_1) = 12$, $N_{F_4}(A_1^2, A_1^2) = 12$, $N_{F_4}(A_1^2, B_2) = 12$, $N_{F_4}(A_2, A_2) = 16$, $N_{F_4}(B_2, B_2) = 3$, $N_{F_4}(A_1^2, A_1, A_1) = 72$, $N_{F_4}(A_2, A_1, A_1) = 48$, $N_{F_4}(B_2, A_1, A_1) = 36$, $N_{F_4}(A_1, A_1, A_1, A_1) = 432$, plus the assignments implied by (45) and (46), all other numbers $N_{F_4}(T_1, \dots, T_d)$ being zero. If one substitutes accordingly in (44), using the information from (48), then one obtains an expression which agrees with (52) after simplification. ◻

17 Proof of the $F = M$ Conjecture for E_6

By (26), we have

$$\begin{aligned}
&(1-xy)^6 F^m_{E_6}\left(\frac{x(1+y)}{1-xy}, \frac{xy}{1-xy}\right)\\
&= \frac{1}{30}m(2m-1)(3m-1)(4m-1)(6m-5)(12m-7)x^6y^6\\
&+ \frac{6}{5}m^2(2m-1)(3m-1)(4m-1)(12m-7)x^6y^5\\
&+ 2m^2(3m-1)(4m-1)\left(36m^2-9m-2\right)x^6y^4\\
&+ 3m^2(4m-1)(4m+1)\left(24m^2-1\right)x^6y^3\\
&+ 2m^2(3m+1)(4m+1)\left(36m^2+9m-2\right)x^6y^2\\
&+ \frac{6}{5}m^2(2m+1)(3m+1)(4m+1)(12m+7)x^6y\\
&+ \frac{1}{30}m(2m+1)(3m+1)(4m+1)(6m+5)(12m+7)x^6\\
&+ \frac{6}{5}m(2m-1)(3m-1)(4m-1)(12m-7)x^5y^5\\
&+ 12m^2(3m-1)(4m-1)(12m-5)x^5y^4 + 6m^2(4m-1)\left(144m^2-12m-5\right)x^5y^3\\
&+ 6m^2(4m+1)\left(144m^2+12m-5\right)x^5y^2 + 12m^2(3m+1)(4m+1)(12m+5)x^5y\\
&+ \frac{6}{5}m(2m+1)(3m+1)(4m+1)(12m+7)x^5 + 2m(3m-1)(4m-1)(30m-13)x^4y^4\\
&+ 6m^2(4m-1)(120m-31)x^4y^3 + 16m^2\left(270m^2-7\right)x^4y^2\\
&+ 6m^2(4m+1)(120m+31)x^4y + 2m(3m+1)(4m+1)(30m+13)x^4\\
&+ 9m(4m-1)(18m-5)x^3y^3 + 18m^2(108m-19)x^3y^2 + 18m^2(108m+19)x^3y\\
&+ 9m(4m+1)(18m+5)x^3 + 12m(21m-4)x^2y^2 + 504m^2x^2y\\
&+ 12m(21m+4)x^2 + 36mxy + 36mx + 1 \qquad (53)
\end{aligned}$$

for the left-hand side of (30).

We now compute the right-hand side of (30) following the proposed procedure in Section 12. The conclusions which we report here are based on *Maple* computations which we performed using Stembridge's `coxeter` package [Ste].

We have $N_{E_6}(E_6) = 1$, $N_{E_6}(A_1 * A_2^2, A_1) = 6$, $N_{E_6}(A_1 * A_4, A_1) = 12$, $N_{E_6}(A_5, A_1) = 6$, $N_{E_6}(D_5, A_1) = 12$, $N_{E_6}(A_1^2 * A_2, A_2) = 36$, $N_{E_6}(A_2^2, A_2) = 8$, $N_{E_6}(A_1 * A_3, A_2) = 24$, $N_{E_6}(A_4, A_2) = 24$, $N_{E_6}(D_4, A_2) = 4$, $N_{E_6}(A_1^2 * A_2, A_1^2) = 18$, $N_{E_6}(A_1 * A_3, A_1^2) = 36$, $N_{E_6}(A_4, A_1^2) = 36$, $N_{E_6}(D_4, A_1^2) = 18$, $N_{E_6}(A_1^3, A_1^3) = 12$, $N_{E_6}(A_1 * A_2, A_1^3) = 24$, $N_{E_6}(A_1 * A_2, A_1 * A_2) = 48$, $N_{E_6}(A_3, A_1^3) = 36$, $N_{E_6}(A_3, A_1 * A_2) = 72$, $N_{E_6}(A_3, A_3) = 27$, $N_{E_6}(A_1^2 * A_2, A_1, A_1) = 144$, $N_{E_6}(A_2^2, A_1, A_1) = 24$, $N_{E_6}(A_1 * A_3, A_1, A_1) = 144$, $N_{E_6}(A_4, A_1, A_1) = 144$, $N_{E_6}(D_4, A_1, A_1) = 48$, $N_{E_6}(A_1^3, A_1^2, A_1) = 180$, $N_{E_6}(A_1^3, A_2, A_1) = 168$, $N_{E_6}(A_1 * A_2, A_1^2, A_1) = 360$, $N_{E_6}(A_1 * A_2, A_2, A_1) = 336$, $N_{E_6}(A_3, A_1^2, A_1) = 378$, $N_{E_6}(A_3, A_2, A_1) = 180$, $N_{E_6}(A_1^2, A_1^2, A_1^2) = 432$, $N_{E_6}(A_2, A_1^2, A_1^2) = 504$, $N_{E_6}(A_2, A_2, A_1^2) = 288$, $N_{E_6}(A_2, A_2, A_2) = 160$,

$N_{E_6}(A_1^2, A_1^2, A_1, A_1) = 2376$, $N_{E_6}(A_2, A_1^2, A_1, A_1) = 1872$, $N_{E_6}(A_2, A_2, A_1, A_1) = 1056$, $N_{E_6}(A_1^3, A_1, A_1, A_1) = 864$, $N_{E_6}(A_1 * A_2, A_1, A_1, A_1) = 1728$, $N_{E_6}(A_3, A_1, A_1, A_1) = 1296$, $N_{E_6}(A_1^2, A_1, A_1, A_1, A_1) = 10368$, $N_{E_6}(A_2, A_1, A_1, A_1, A_1) = 6912$, $N_{E_6}(A_1, A_1, A_1, A_1, A_1, A_1) = 41472$, plus the assignments implied by (45) and (46), all other numbers $N_{E_6}(T_1, \dots, T_d)$ being zero. If one substitutes accordingly in (44), using the information from (48), then one obtains an expression which agrees with (53) after simplification. □

18 The Dual F-triangle

Armstrong [Arm06+, Sec. 4] defines the *dual F-triangle*, denoted here by $\tilde{F}_\Phi^m(x, y)$, as

$$\tilde{F}_\Phi^m(x, y) = (-1)^n F_\Phi^m(-1 - x, -1 - y),$$

where n is the rank of the root system Φ. He conjectures that the dual F-triangle can be expressed in form of a weighted bivariate generating function for the faces of $\Delta^m(\Phi)$ involving the *Fuss–Narayana numbers* $\mathrm{Nar}^m(\Phi, i)$, the latter enumerating all elements of rank i in the m-divisible non-crossing partition poset $NC^m(\Phi)$. For explicit formulae for the Fuss–Narayana numbers see [Arm06+, Sec. 2]. These numbers occur also as h-numbers in [FR05, Theorem 9.2]. (One has to reverse the ordering of the numbers to convert one sequence of numbers into the other.) In view of our proof below, Armstrong's conjecture becomes the following theorem.

Theorem DF. *For any finite root system Φ, we have*

$$\tilde{F}_\Phi^m(x, y) = \sum_{k,l \ge 0} \frac{\mathrm{Nar}^m(\Phi, k + l)}{\mathrm{Nar}^1(\Phi, k + l)} f_{k,l}\, x^k y^l. \tag{54}$$

Proof. Clearly, for the exceptional root systems one can verify (54) routinely by using the explicit formulae for the refined face numbers, as given through the formulae for the F-triangle in Section 7, and the formulae for the Fuss–Narayana numbers in [Arm06+, FR05].

To verify (54) for the root systems A_n, B_n and D_n, some work has to be done. However, the verifications in these types are very similar to each other so that we give below only the proof in type A_n, leaving the proofs for B_n and D_n to the reader.

By Theorem FA, in type A_n the left-hand side of (54) is equal to

$$(-1)^n \sum_{k,l \ge 0} \frac{l+1}{k+l+1} \binom{n}{k+l} \binom{m(n+1)+k-1}{k} (-1-x)^k (-1-y)^l$$
$$= \sum_{k,l,r,s \ge 0} \frac{l+1}{k+l+1} \binom{n}{k+l} \binom{m(n+1)+k-1}{k} \binom{k}{r} \binom{l}{s} (-1)^{n+k+l} x^r y^s$$

$$= \sum_{k,r,s\ge 0} \frac{s+1}{n+1} \binom{n-s-1}{n-k-s} \binom{m(n+1)+k-1}{k} \binom{k}{r} (-1)^{n+k+s} x^r y^s$$

$$= \sum_{k,r,s\ge 0} \frac{s+1}{n+1} \binom{n-s-1}{n-k-s} \binom{m(n+1)+k-1}{k-r} \binom{m(n+1)+r-1}{r} (-1)^{n+k+s} x^r y^s$$

$$= \sum_{r,s\ge 0} \frac{s+1}{n+1} \binom{m(n+1)}{n-s-r} \binom{m(n+1)+r-1}{r} x^r y^s.$$

As earlier, for the evaluation of the sums over l and k we used special instances of the Chu–Vandermonde summation.

On the other hand, by Theorem FA and by [Arm06$^+$, Sec. 2], the right-hand side of (54) is equal to

$$\sum_{k,l\ge 0} \frac{\frac{1}{n+1}\binom{n+1}{k+l}\binom{m(n+1)}{n-k-l}}{\frac{1}{n+1}\binom{n+1}{k+l}\binom{n+1}{n-k-l}} \frac{l+1}{k+l+1} \binom{n}{k+l} \binom{m(n+1)+k-1}{k} x^k y^l,$$

which is exactly the same expression. □

Acknowledgement. I am indebted to Sergey Fomin for suggesting to me to embark on this project and, thus, to the Institut Mittag–Leffler, Anders Björner and Richard Stanley for hosting both of us at the same time at the Institut during the "Algebraic Combinatorics" programme in Spring 2005. Moreover, I wish to thank Drew Armstrong, Christos Athanasiadis and Eleni Tzanaki for the many extremely helpful exchanges that I had with them on the subject matter, and for making drafts for [Arm06$^+$], [Ath06$^+$] and [Tza06$^+$] available to me at an early stage. Last, but not least, I am indebted to the referee for an extremely careful reading of the manuscript.

References

[Arm06$^+$] D. Armstrong, Generalized noncrossing partitions and combinatorics of Coxeter groups, preprint.

[Ath06$^+$] C. A. Athanasiadis, On some enumerative aspects of generalized associahedra, preprint; `arXiv:math.CO/0508030`.

[AR04] C. A. Athanasiadis and V. Reiner, Noncrossing partitions for the group D_n, *SIAM J. Discrete Math.* **18** (2004), 397–417.

[Bes03] D. Bessis, The dual braid monoid, *Ann. Sci. École Norm. Sup.* (4) **36** (2003), 647–683.

[BW02] T. Brady and C. Watt, $K(\pi, 1)$'s for Artin groups of finite type, *Geom. Dedicata* **94** (2002), 225–250.

[BW06$^+$] T. Brady and C. Watt, Lattices in finite reflection groups, preprint; `arXiv:math.CO/0501502`.

[Car77] L. Carlitz, Some expansion and convolution formulas related to MacMahon's master theorem, *SIAM J. Math. Anal.* **8** (1977), 320–336.

[Cha04] F. Chapoton, Enumerative properties of generalized associahedra, *Séminaire Lotharingien Combin.* **51** (2004), Article B51b, 16 pp.

[Ede81] P. Edelman, Chain enumeration and noncrossing partitions, *Discrete Math.* **31** (1981), 171–180.

[FR05] S. Fomin and N. Reading, Generalized cluster complexes and Coxeter combinatorics, *Int. Math. Res. Notices* **44** (2005), 2709–2757.

[FZ02] S. Fomin and A. Zelevinsky, Cluster algebras I: Foundations, *J. Amer. Math. Soc.* **15** (2002), 497–529.

[FZ03a] S. Fomin and A. Zelevinsky, Cluster algebras II: Finite type classification, *Invent. Math.* **154** (2003), 63–121.

[FZ03b] S. Fomin and A. Zelevinsky, Y-systems and generalized associahedra, *Ann. of Math.* (2) **158** (2003), 977–1018.

[GKP89] R. L. Graham, D. E. Knuth and O. Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, Massachusetts, 1989.

[Hum90] J. E. Humphreys, *Reflection groups and Coxeter groups*, Cambridge University Press, Cambridge, 1990.

[Kre72] G. Kreweras, Sur les partitions non croisées d'un cycle, *Discrete Math.* **1** (1972), 333–350.

[Rei97] V. Reiner, Non-crossing partitions for classical reflection groups, *Discrete Math.* **177** (1997), 195–222.

[Sla66] L. J. Slater, *Generalized hypergeometric functions*, Cambridge University Press, Cambridge, 1966.

[Sta98] R. P. Stanley, *Enumerative Combinatorics*, Vol. 1, Wadsworth & Brooks/Cole, Pacific Grove, California, 1986; reprinted by Cambridge University Press, Cambridge, 1998.

[Ste] J. R. Stembridge, *coxeter*, *Maple* package for working with root systems and finite Coxeter groups; available at `http://www.math.lsa.umich.edu/~jrs`.

[Tza06+] E. Tzanaki, *Combinatorics of generalized cluster complexes and hyperplane arrangements*, Ph.D. thesis, University of Crete, Iraklio, in preparation.

Part II

Ramsey Theory

Monochromatic Equilateral Right Triangles on the Integer Grid

Ron Graham[1⋆] and József Solymosi[2⋆⋆]

[1] Department of Computer Science & Engineering, University of California, San Diego, USA
graham@ucsd.edu
[2] Department of Mathematics, University of British Columbia, Vancouver, Canada
solymosi@math.ubc.ca

Summary. For any coloring of the $N \times N$ grid using fewer than $\log\log N$ colors, one can always find a **monochromatic** equilateral right triangle, a triangle with vertex coordinates $(x, y), (x + d, y)$, and $(x, y + d)$.

AMS Subject Classification. 05D10.
Keywords. Van der Waerden, Gallai–Witt Theorem.

1 Introduction

The celebrated theorem of van der Waerden [Wae27] states that for any natural numbers k and r, there is a number $W(k, r)$ such that for any coloring of the first $W(k, r)$ natural numbers by r colors, there is always a monochromatic arithmetic progression of length k. Answering a question of Erdős and Turán [ET36], Roth [Roth53] proved a density version of van der Waerden's theorem for $k = 3$. He proved that $r_3(N)$, the cardinality of the largest subset of $\{1, \ldots, N\}$ containing no three distinct elements $x, x + d, x + 2d$ in arithmetic progression, is $O(N/\log\log N)$. This was not only the first proof for the conjecture of Erdős and Turán, but also the first efficient bound on $W(3, r)$. One of the goals of the present paper is to give a combinatorial proof of such a bound, proving that $W(3, r) \leq 2^{2^{cr}}$. The best known bound for $W(3, r)$ is the one which follows from Bourgain's [Bou99] result $r_3(N) = O(N(\log\log N/\log N)^{1/2})$, which is better than ours, but uses heavy tools from analysis. Van der Waerden's Theorem was extended by Gallai, proving that in any finite coloring of $\mathbb{Z}^2$, some color contains arbitrarily large square subarrays. The simplest density version of this extension is to prove

⋆ Research supported in part by Grant CCR-0310991.
⋆⋆ Research supported by NSERC and OTKA grants.

that there is always a triangle in a dense $N \times N$ grid with vertex coordinates $(x, y), (x + d, y)$, and $(x, y + d)$, if N is large enough compared to the density. This was first asked by Erdős and Graham in [EG80]. The first proof of the statement was given by Ajtai and Szemerédi [AS74] and later a much more general theorem, the so called Multidimensional Szemerédi Theorem [Sze75] was presented by Fürstenberg and Katznelson [FK79]. The proofs gave no (or very weak) bounds for the maximum density of subsets of the grid avoiding such triangles. The best bound is due to Shkredov [Shk], who proved that if the density of a subset of the $N \times N$ grid is at least $1/(\log\log\log N)^c$ then it contains a triangle. Our main result is the following.

Theorem 1. *There is a universal $c > 0$, such that for any coloring of the $N \times N$ grid by no more than $c \log\log N$ colors, there is always a monochromatic triangle with vertex coordinates $(x, y), (x + d, y)$, and $(x, y + d)$.*

Corollary 2 (van der Waerden's Theorem, $k = 3$ case). *For any coloring of $[N]$ by no more than $c \log\log N$ colors, there is always a monochromatic arithmetic progression of length 3. Using the usual notation, $W(3, k) \leq 2^{2^{ck}}$.*

Proof. Every coloring of the set $\mathbb{Z}$ of integers defines a coloring of $\mathbb{Z}^2$ by giving the color of $x - y$ to the point with coordinates (x, y). In this way, a monochromatic triple with vertex coordinates $(x, y), (x + d, y)$, and $(x, y + d)$, defines a monochromatic arithmetic progression $x - y - d, x - y, x - y + d$. □

It is worth mentioning that the traditional combinatorial proof using color focusing gives

$$W(3, k) \leq \left. k^{k^{k^{\cdot^{\cdot^{\cdot^{k^{4k}}}}}}} \right\} (k - 1),$$

a tower-type bound.

2 Proof of Theorem 1

Let us suppose that the points of the $N \times N$ grid are colored by L colors, and there is no monochromatic equilateral right triangle. We will show that L must be large. Let us examine the coloring of the elements of the points on the diagonal of the grid, i.e., the points with coordinates (x, y) such that $x + y = N + 1$. Select the most popular color, denoted by c_1. The set of points of the diagonal with color c_1 is denoted by S_1. For any pair $p = (a, b), q = (c, d)$, elements of S_1, the points (a, d) and (c, b) cannot have the color c_1. The Cartesian product defined by the points of S_1 has the property that only the diagonal has points with color c_1. The lower-triangular part of the Cartesian product is denoted by T_1, i.e.,

$$T_1 = \{(x, y) : \exists\, s, t \ni (x, t), (s, y) \in S_1, s > x\}$$

Note that $s_1 := |S_1| \geq \frac{N}{L}$. We now define the color c_{i+1}, the set S_{i+1}, and T_{i+1} recursively, based on c_i, S_i, and T_i (where $i \geq 1$).

Suppose the pointset T_i avoids the colors $c_1, c_2, \ldots, c_i$. There is a line with slope -1, which contains many points of T_i. Let m be such that

$$|\{(x,y) : x+y=m\} \cap T_i| \geq \frac{|T_i|}{N}.$$

Select the points with the most popular color, c_{i+1}, in T_i along the line $x+y=m$. The set of these points will be S_{i+1}, and

$$T_{i+1} = \{(x,y) : \exists\, s,t \ni (x,t),(s,y) \in S_{i+1}, s > x\}.$$

Thus, the pointset T_{i+1} avoids the colors $c_1, c_2, \ldots, c_{i+1}$. Note that we have the inequality

$$s_{i+1} = |S_{i+1}| \geq \frac{\binom{s_i}{2}}{(L-i)N}.$$

If we reach Step L with $s_L \geq 2$ then we have a contradiction, since we run out of colors for T_L.

From the formula above, one can already get a feeling for the magnitude of the bound. However, for the formal proof of Theorem 1, we prove the following.

Lemma 3. *If $s_1 \geq N/r$, $s_{i+1} \geq \frac{1}{(r-i)N}\binom{s_i}{2}$ and $N = N(r) = (2r)^{2^r}$ then $s_r \geq 2$.*

Proof. We prove by induction on i that for $1 \leq i \leq r$, we have:

(a) $s_i \geq \dfrac{N}{2^{2^{i-1}-1} r^{2^i-1}}$,

(b) $s_i \geq r/i$.

This is clearly true for $i=1$. Suppose it is true for some $i < r$. Then

$$s_{i+1} \geq \frac{1}{(r-i)N}\binom{s_i}{2} = \frac{s_i^2}{2rN} \cdot \frac{r}{r-i} \cdot \frac{s_i - 1}{s_i}$$

But

$$\frac{r}{r-i} \cdot \frac{s_i - 1}{s_i} \geq 1$$

since $s_i \geq r/i$ by induction. Hence, we have

$$s_{i+1} \geq \frac{s_i^2}{2rN} \geq \frac{1}{2rN} \cdot \frac{N^2}{2^{2^i-2} r^{2^{i+1}-2}} = \frac{N}{2^{2^i-1} r^{2^{i+1}-1}}$$

which is (a) for $i+1$. It is easy to see that (b) also holds for $i+1$ as well. The inequality for s_r is now

$$s_r \geq \frac{(2r)^{2^r}}{2^{2^{r-1}-1} r^{2^r-1}} \geq 2^{2^{r-1}+1} r \geq 2.$$

This completes the proof of the lemma and Theorem 1. □

We note here that with a similar but somewhat more complicated argument, we can prove that there are many monochromatic corners when the number of colors is small. In particular, we can show:

Theorem 4. *For any integer $r > 0$, if the lattice points in the $N \times N$ grid are arbitrarily r-colored, and $N > 2^{2^{3^r}}$ then there are always at least $\delta(r)N^3$ monochromatic "corners", i.e., triples of points $(x, y), (x + d, y), (x, y + d)$ for some $d > 0$, where $\delta(r) = (3r)^{-2^{r+2}}$.*

We note that this is similar in spirit to the results of [FGR88] where it is shown that in fact a **positive fraction** of the objects being colored must occur monochromatically. The proof follows that of Theorem 1 and is omitted.

We should also point out that this approach can be used to prove directly a quantitative version van der Waerden's theorem for 3-term arithmetic progressions, namely that if $\mathbf{Z}_p$ is colored by at most $c \log\log p$ colors, then some monochromatic 3-term arithmetic progression must be formed. Similarly, analogous results can be obtained for the occurrence of monochromatic affine lines in $GF(3)^n$ using this approach.

Acknowledgement. The authors wish to acknowledge the insightful comments of Fan Chung and Steven Butler in preparing the final version of the manuscript.

References

[AS74] M. Ajtai and E. Szemerédi, Sets of lattice points that form no squares, *Stud. Sci. Math. Hungar.* **9** (1974), 9–11.

[Bou99] J. Bourgain, On triples in arithmetic progression. *Geom. Func. Anal.* **9** (1999), 968–984.

[EG80] P. Erdős and R. L. Graham, 'Old and new problems and results in combinatorial number theory', *Monographies de L'Enseignement Mathématique*, vol 28, Université de Genève L'Enseignement Mathématique, (1980).

[ET36] P. Erdős and P. Turán, On some sequences of integers, *J. London Math. Soc.* **11** (1936), 261–264.

[FGR88] P. Frankl, R.L. Graham and V. Rödl, Quantitative theorems for regular systems of equations, *Jour. Comb. Th. A* **47** (1988), 246–261.

[FK79] H. Fürstenberg, and Y. Katznelson, An ergodic Szemerédi Theorem for commuting transformations, *J. Analyse Math.* **34** (1979), 275–291.

[Roth53] K.F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.

[Shk] I.D. Shkredov, On one problem of Gowers, Math. Arxiv math. NT/0405406.

[Sze75] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245.

[Wae27] B. L. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wiskunde* **15** (1927), 212–216.

One-sided Coverings of Colored Complete Bipartite Graphs

András Gyárfás[1], Miklós Ruszinkó[1⋆], Gábor N. Sárközy[2⋆⋆], and Endre Szemerédi[3]

[1] Computer and Automation Research Institute, Hungarian Academy of Sciences, P.O. Box 63, Budapest, H-1518 Hungary
gyarfas,ruszinko@sztaki.hu

[2] Computer Science Department, Worcester Polytechnic Institute, Worcester, MA, 01609 USA, and Computer and Automation Research Institute, Hungarian Academy of Sciences, P.O. Box 63, Budapest, H-1518 Hungary
gsarkozy@cs.wpi.edu

[3] Computer Science Department, Rutgers University New Brunswick, NJ, 08903 USA
szemered@cs.rutgers.edu

Summary. Assume that the edges of a complete bipartite graph $K(A, B)$ are colored with r colors. In this paper we study coverings of B by vertex disjoint monochromatic cycles, connected matchings, and connected subgraphs. These problems occur in several applications.

AMS Subject Classification. 05C55, 05C38.
Keywords. Vertex coverings by monochromatic subgraphs, Regularity Lemma.

1 Introduction

Some problems for edge colored complete graphs naturally lead to edge colored complete bipartite graphs. For example, in [Gya77] it was proved that in every r-coloring of the edges of K_n there is a connected monochromatic subgraph of order at least $\frac{n}{r-1}$. The proof was based on the result that in every $(r-1)$-coloring of the edges of a complete bipartite graph of order n there is a connected monochromatic subgraph of order at least $\frac{n}{r-1}$. (We remark here that later Füredi [Für81] obtained an important result on fractional matchings of hypergraphs which also implies the cited result.) As another example, in [EGP91] it was proved that the vertex set of an r-colored complete

⋆ Research supported in part by OTKA Grants T038198, T046234 and
⋆⋆ by the National Science Foundation under Grant No. DMS-0456401.

graph can be covered by at most $cr^2 \log r$ vertex disjoint monochromatic cycles. (Throughout this paper an edge or a single vertex is considered to be a cycle, too.) The proof used the following "one-sided" covering lemma for bipartite graphs. If $G = K(A, B)$ is an r-colored complete bipartite graph with $|A| \geq r^3|B|$ then B can be covered by the vertices of at most r^2 vertex disjoint monochromatic cycles. This lemma was strengthened in [GRSSz06] by showing that at most $(6r\lceil \log r\rceil + 2r)$ vertex disjoint monochromatic cycles suffice to cover B if $|A| \geq r^2|B|$. This result has been used by the same set of authors to improve the result cited above as follows.

Theorem 1.1 (Gyárfás, Ruszinkó, Sárközy, Szemerédi [GRSSz06]). *For every integer $r \geq 2$ there exists a constant $n_0 = n_0(r)$ such that if $n \geq n_0$ and the edges of the complete graph K_n are colored with r colors then the vertex set of K_n can be partitioned into at most $100r \log r$ vertex disjoint monochromatic cycles.*

In these improvements, as in this paper, the Regularity Lemma played a major role.

In this paper one-sided coverings of colored complete bipartite graphs are explored further. The main result is the following improved form of the one-sided covering lemma for cycles.

Theorem 1.2. *For every fixed r there exists $n_0 = n_0(r)$ such that the following is true. Assume that the edges of a complete bipartite graph $K(A, B)$ are colored with r colors, where $|A| \geq n_0$. If $|A| \geq 2r|B|$, then B can be covered by at most $3r$ vertex disjoint monochromatic cycles.*

Note that this is a significant improvement over the above cited result from [GRSSz06], where the statement is proved with $(6r\lceil \log r\rceil + 2r)$ cycles instead of $3r$ cycles for $|A| \geq r^2|B|$. However, this result in itself does not give an improvement in Theorem 1.1. Indeed, the magnitude $r \log r$ occurs in two different steps of the proof: In the "ancestor" of Theorem 1.2 (with $6r\lceil \log r\rceil + 2r$) and in some other iterative step. Unfortunately, we could not give a linear bound in the latter one but this is the only place at this stage with a non-linear bound. Thus Theorem 1.2 is a significant step toward a linear bound in Theorem 1.1 – if such exists at all.

One tool of the proof, interesting in its own, is Theorem 2.1 which has an easy elementary proof. It says that the condition $|A| \geq r|B|$ ensures that in an r-colored complete bipartite graph $K(A, B)$, B can be covered by at most r vertex disjoint monochromatic connected matchings, in fact one can require that each matching has a distinct color. Here a monochromatic (say red) connected matching is a matching that lies in the same red connected component. Note that monochromatic connected matchings also played an important role in [Łuc99], [GRSSz06]. Luczak [Łuc99] realized (through the Regularity Lemma) that the Ramsey numbers of monochromatic connected matchings and paths are about the same. Using this method the same set of authors [GRSSz06] determined exactly the three color Ramsey numbers for paths which was an open problem for more than twenty years.

Theorem 2.1 is close to best possible: there are infinitely many r-colored complete bipartite graphs $K(A_m, B_m)$ such that $|A| = |B|(r-1-\frac{r-1}{mr!})$ and B cannot be covered by the vertices of at most r vertex disjoint connected monochromatic matchings (Corollary 2.6).

We also prove that the (much) weaker condition $|B| < e^{|A|/r^{r+3}} - |A|$ is enough to ensure a covering of B with at most r vertex disjoint monochromatic connected subgraphs (Corollary 2.12). This result is obtained through Theorem 2.11, a generalization of a result of Haxell and Kohayakawa ([HK96]). Notice that for $|B| \geq r$ one can color $K(A, B)$ by defining a partition of B into r nonempty parts and color all edges between A and the i-th part by color i. This coloring shows that in one sided coverings of complete bipartite graphs at least r monochromatic subgraphs are needed.

2 One-sided Covers of Bipartite Graphs

In certain covering or partition problems one may require that all monochromatic objects have distinct colors, i.e. color repetition is not allowed. For example, it is not known whether every 3-colored complete graph can be covered by three monochromatic paths but there are examples when there is no cover if we want paths of distinct colors. Another example is the result of Haxell and Kohayakawa proving that every r-colored complete graph can be partitioned into at most r monochromatic trees of distinct colors. In this section we prove two lemmas about one-sided coverings where the colors of the objects are all different.

2.1 Covering B by Monochromatic Connected Matchings

Theorem 2.1. *Assume that the edges of a complete bipartite graph $K(A, B)$ are colored with r colors, $|A| \geq r|B|$. Then there are vertex disjoint monochromatic connected matchings, all of different color, such that their union covers each vertex of B.*

Proof. We define by iteration r-colored complete bipartite graphs $G_i = K(A \setminus A_i, B)$, $\overline{G_i} = K(A_i, B)$ and sets $X_i \subseteq A_i, Y_i \subseteq B$, such that $A_i = \cup_{j=0}^{i} X_j$. Initially $G_0 = G, A_0 = X_0 = Y_0 = \emptyset$.

The general step is to select an arbitrary vertex $a \in A \setminus A_{i-1}$ and consider the partition $\mathcal{P}$ of B by putting two vertices $p, q \in B$ into the same class if and only if the colors of ap, aq are the same and label the class by the color of ap. Let E be defined as the set of those edges $a'b$ of G_{i-1} whose color is the same as the label of the class of $\mathcal{P}$ containing b. Observe that the existence of a matching of B to A_{i-1} using edges of E proves the theorem - then the procedure stops. Therefore we may assume that such a matching does not exist. By Hall's theorem there are sets $X_i \subseteq A \setminus A_{i-1}, Y_i \subseteq B$ such that $|X_i| < |Y_i|$ and all edges of E incident to Y_i are incident to X_i (i.e. X_i is the set of E-neighbors of Y_i). Set $A_i = A_{i-1} \cup X_i$ and let G_i be the complete bipartite subgraph of G spanned by $[A \setminus A_i, B]$. Notice that $a \in X_i$ thus at least one new vertex is added to A_i. This finishes the definitions for step i.

Since at each step $|A_i| > |A_{i-1}|$, the procedure terminates with $A_m = A$ (and $G_m = \emptyset$) for some m. We show that this leads to a contradiction, thus the procedure must terminate with finding the required cover of B.

Assume that some vertex $b \in B$ in $\overline{G_m}$ is covered by $Y_1, Y_2, \dots, Y_k$. Then there are k distinct colors such that all edges incident to b in one of these colors go to $\cup_{i=1}^{k} X_i$. Therefore b is incident to edges of at most $r-k$ colors in G_k implying $k \leq r$. Assuming that the procedure takes m steps, consider the hypergraph on vertex set B with edges Y_i,

$$r|B| \geq \sum_{x \in B} d(x) = \sum_{j=1}^{m} |Y_j| \geq \sum_{j=1}^{m} (|X_j| + 1) = |A| + m > |A| \qquad (1)$$

contradicting the assumption of the theorem. □

A bipartite graph $G(k,l)$ is **γ-dense** if it contains at least γkl edges. We will need the following $(1-\varepsilon)$-dense version of Theorem 2.1 as well.

Theorem 2.2. *For some $0 < \varepsilon < 1/9$ assume that the edges of a $(1-\varepsilon)$-dense bipartite graph $G(A,B)$ are colored with r colors, $|A| \geq 3r|B|/2$. Then there are vertex disjoint monochromatic connected matchings, each of a different color, such that their union covers at least $(1-\sqrt{\varepsilon})$-fraction of the vertices of B.*

Proof. First we "trim" $G(A,B)$, we keep only the high degree vertices. For this purpose we use the following fact.

Fact 2.3. *Let $G(A,B)$ be a $(1-\varepsilon)$-dense bipartite graph. Then there is a subset $A' \subseteq A$ with $|A'| \geq (1-\sqrt{\varepsilon})|A|$ such that $\deg(a,B) \geq (1-\sqrt{\varepsilon})|B|$ for all $a \in A'$.*

Indeed we get A' by removing those vertices from A that have degree less than $(1-\sqrt{\varepsilon})|B|$ in G. The number of these vertices is at most $\sqrt{\varepsilon}|A|$ from the density condition.

The proof of Theorem 2.2 is similar to that of Theorem 2.1, but we always select the vertex a from the set A' and then we get a partition $\mathcal{P}$ of the $(\geq (1-\sqrt{\varepsilon})|B|)$ neighbors of a in B. Then a matching covering these neighbors gives the desired covering with monochromatic connected matchings. The contradiction in (1) is similar, since we stop if there are no more A' vertices in the leftover:

$$r|B| \geq \dots > (1-\sqrt{\varepsilon})|A|,$$

which is a contradiction for $0 < \varepsilon < 1/9$ and $|A| \geq 3r|B|/2$. □

To see that Theorem 2.1 cannot be improved too much, let $G_1 = K(A, B)$ be the following r-colored complete bipartite graph. Set $A = [r]$ and each vertex of B is associated with a permutation of $[r]$. Vertex $i \in A$ is adjacent to a permutation in B in the color which is the i-th element of the permutation.

Lemma 2.4. *Assume that $\mathcal{S} = \{S_1, S_2, \ldots, S_t\}$ are monochromatic stars of G_1 with their centers in A and such that the union of their leaves cover B. Then $t \geq r$ with equality if and only if : (i) all centers coincide and all colors are different, or (ii) all centers are different and all colors are the same.*

Proof. Suppose that $X_i \subseteq [r]$, $i \in [r]$, is the set of colors (we always color by colors $1, 2, \ldots, r$) appearing on the members of $\mathcal{S}$ with center at $i \in A$. The sets $\overline{X_i} = [r] \setminus X_i$ have no distinct representatives. Indeed, the existence of such a set of representatives is equivalent to the existence of a vertex of B uncovered by the leaves of the stars, contradicting the assumption. Thus, by Hall's theorem, there exists a set $A^* \subseteq A$ such that $|A^*| = j$ and $|\cup_{i \in A^*} \overline{X_i}| \leq j - 1$ implying that $|\cap_{i \in A^*} X_i| \geq r - j + 1$. Therefore

$$t \geq \sum_{i \in A^*} |X_i| \geq |A^*| |\cap_{i \in A^*} X_i| \geq j(r - j + 1) \geq r$$

with equality in the last inequality if and only if $j = 1$ or $j = r$ giving cases (i) and (ii) in the lemma. □

The following corollary shows that r cannot be essentially lowered in the condition $|A| \geq r|B|$ of Theorem 2.1.

Corollary 2.5. *For every fixed r there are infinitely many r-colored complete bipartite graphs $[A_m, B_m]$ such that $|A_m| = |B_m|(r - \frac{1}{m(r-1)!})$ and B_m cannot be covered by the vertices of vertex disjoint connected monochromatic matchings, each having a different color.*

Proof. Consider the graph $G_1 = K(A, B)$ and replace each vertex of B by a set of m vertices, each vertex of A by a set of $mr! - 1$ vertices. This gives an r-colored complete bipartite graph $G_1^m = K(A_m, B_m)$ with $|B_m| = mr!$, $|A_m| = r(mr! - 1) = |B_m|(r - \frac{1}{m(r-1)!})$ for every positive integer m. Since for any $x \in B$ two edges of G_1 incident to x are always colored with different color, a connected monochromatic matching in G_m corresponds (can be contracted) to a monochromatic star in G_1 with center in A. Thus the required covering of B_m with disjoint monochromatic matchings corresponds to a star-cover as in Lemma 2.4. Applying Lemma 2.4, the only possibility to cover B_m is coming from (i), i.e. all monochromatic matchings are using the vertices of a replacement of a single vertex of $A \subseteq V(G_1)$. Since any vertex of A is replaced by $mr! - 1$ vertices there is no matching from that set to B_m since $|B_m| = mr!$. □

If one does not require that all monochromatic connected matchings have distinct colors we have only a weaker construction:

Corollary 2.6. *For every fixed r there are infinitely many r-colored complete bipartite graphs $K(A_m, B_m)$ such that $|A_m| = |B_m|(r - 1 - \frac{r-1}{mr!})$ and B_m cannot be covered by the vertices of at most r vertex disjoint connected monochromatic matchings.*

Proof. It is similar to the proof of Corollary 2.5. The only difference is that here we use G_1^* obtained from G_1 by deleting an arbitrary vertex of A. Then, using the same replacements as in the proof of Corollary 2.5, possibility (*ii*) of an r-covering is eliminated from Lemma 2.4 and the proof follows. □

2.2 Covering B by Monochromatic Cycles

In this section we prove our main result, Theorem 1.2. We will use the bipartite r-color version of the Regularity Lemma (for an extensive survey on different variants of the Regularity Lemma see [KS96]). For this purpose we will need some definitions. For non-empty A and B,

$$d_G(A, B) = \frac{e_G(A, B)}{|A||B|}$$

is the **density** of the graph between A and B.

Definition 2.7. *The bipartite graph $G = (A, B, E)$ is* $(\boldsymbol{\varepsilon}, \boldsymbol{G})$**-regular** *if*

$$X \subset A,\ Y \subset B,\ |X| > \varepsilon|A|,\ |Y| > \varepsilon|B| \text{ imply } |d_G(X, Y) - d_G(A, B)| < \varepsilon,$$

otherwise it is $(\boldsymbol{\varepsilon}, \boldsymbol{G})$**-irregular**. *Furthermore,* (A, B, E) *is* $(\boldsymbol{\varepsilon}, \boldsymbol{\delta}, \boldsymbol{G})$**-super-regular** *if it is* (ε, G)*-regular and*

$$\deg_G(a) > \delta|B| \ \forall\, a \in A, \qquad \deg_G(b) > \delta|A| \ \forall\, b \in B.$$

Proof of Theorem 1.2. Consider an r-edge coloring $(G_1, G_2, \ldots, G_r)$ of $K(A, B)$. We apply the bipartite r-color version of the Regularity Lemma with a sufficiently small ε. By standard arguments we may assume that for each cluster that is not V_0, all vertices of the cluster belong to the same partite class. Thus we get a partition $A = V_A^0 + V_A^1 + \ldots V_A^{l_A}$, $B = V_B^0 + V_B^1 + \ldots V_B^{l_B}$, where $|V_A^{j_1}| = |V_B^{j_2}| = m$, $1 \le j_1 \le l_A$, $1 \le j_2 \le l_B$ and $|V_A^0| \le \varepsilon|A|$, $|V_B^0| \le \varepsilon|B|$. We define the reduced graph G^R: The vertices of G^R are $A^R = \{p_A^{j_1} \mid 1 \le j_1 \le l_A\}$ and $B^R = \{p_B^{j_2} \mid 1 \le j_2 \le l_B\}$, and we have an edge between vertices $p_A^{j_1}$ and $p_B^{j_2}$, if the pair $\{V_A^{j_1}, V_B^{j_2}\}$ is (ε, G_s)-regular for $s = 1, 2, \ldots, r$. Thus we have a one-to-one correspondence $f : \{p_A^j, p_B^j\} \to \{V_A^j, V_B^j\}$ between the vertices of G^R and the non-exceptional clusters of the partition. Then $G^R = (A^R, B^R)$ is a $(1 - \varepsilon)$-dense bipartite graph. Define an r-edge coloring $(G_1^R, G_2^R, \ldots, G_r^R)$ of G^R in the following way. The edge between the clusters $V_A^{j_1}$ and $V_B^{j_2}$ is colored with a color s that contains the most edges from $K(V_A^{j_1}, V_B^{j_2})$, thus clearly

$$\left|E_{G_s}(V_A^{j_1}, V_B^{j_2})\right| \geq \frac{1}{r}|V_A^{j_1}||V_B^{j_2}|.$$

Applying Theorem 2.2 to G^R we get at most r vertex disjoint monochromatic connected matchings that cover at least $(1-\sqrt{\varepsilon})$-fraction of the vertices of B^R. The clusters not covered by these monochromatic connected matchings are placed into the exceptional set V_B^0. With standard techniques, going back to the original graph, from these monochromatic connected matchings we can construct monochromatic cycles that cover most of the clusters belonging to these connected matchings. Indeed, let us take a monochromatic connected matching M, say M is in G_1^R and has size $|M| = l_1$. We will make this connected matching into a cycle in G_1.

Denote the matching $M = \{e_1, e_2, \ldots, e_{l_1}\}$ between the two sets of end points $U_A \subseteq A^R$ and $U_B \subseteq B^R$. Furthermore, let $f(e_i) = (V_A^i, V_B^i)$ for $1 \leq i \leq l_1$ where V_A^i and V_B^i are the clusters assigned to the endpoints of e_i.

We need to do some preparations on the matching M. First we will find connecting paths between the edges of the matching M. Since M is a connected matching in G_1^R we can find l_1 connecting paths P_i^R in G_1^R from $f^{-1}(V_B^i)$ to $f^{-1}(V_A^{i+1})$ for every $1 \leq i \leq l_1$ (for $i = l_1$ we go from $f^{-1}(V_B^{l_1})$ back to $f^{-1}(V_A^1)$). Note that these paths in G_1^R may not be internally vertex disjoint. From these paths P_i^R in G_1^R we can construct vertex disjoint connecting paths P_i in G_1 connecting a typical vertex v_B^i of V_B^i to a typical vertex v_A^{i+1} of V_A^{i+1}. More precisely we construct P_1 with the following simple greedy strategy. Denote $P_1^R = (p_1, \ldots, p_t), 2 \leq t \leq l_A + l_B$, where according to the definition $f(p_1) = V_B^1$ and $f(p_t) = V_A^2$. Let the first vertex u_1 $(= v_B^1)$ of P_1 be a vertex $u_1 \in V_B^1$ for which $\deg_{G_1}(u_1, f(p_2)) \geq (1/r - \varepsilon)m$ and $\deg_{G_1}(u_1, V_A^1) \geq (1/r - \varepsilon)m$. By (ε, G_1)-regularity most of the vertices satisfy this in V_B^1. The second vertex u_2 of P_1 is a vertex $u_2 \in (f(p_2) \cap N_{G_1}(u_1))$ for which $\deg_{G_1}(u_2, f(p_3)) \geq (1/r - \varepsilon)m$. Again by (ε, G_1)-regularity most vertices satisfy this in $f(p_2) \cap N_{G_1}(u_1)$. The third vertex u_3 of P_1 is a vertex $u_3 \in (f(p_3) \cap N_{G_1}(u_2))$ for which $\deg_{G_1}(u_3, f(p_4)) \geq (1/r - \varepsilon)m$. We continue in this fashion, finally the last vertex u_t $(= v_A^2)$ of P_1 is a vertex $u_t \in (f(p_t) \cap N_{G_1}(u_{t-1}))$ for which $\deg_{G_1}(u_t, V_B^2) \geq (1/r - \varepsilon)m$.

Then we move on to the next connecting path P_2. Here we follow the same greedy procedure, we pick the next vertex from the next cluster in P_2^R. However, if the cluster has occurred already on the path P_1^R (or on any other connecting paths later in the procedure), then we just have to make sure that we pick a vertex that has not been used so far. Since the total number of vertices on the connecting paths will be a constant, this is feasible.

We continue in this fashion and construct the vertex disjoint connecting paths P_i in G_1, $1 \leq i \leq l_1$. These will be parts of the final cycle in G_1. We remove the internal vertices of these paths from G_1. Furthermore, we remove some more vertices from each $(V_A^i, V_B^i), 1 \leq i \leq l_1$ to achieve super-regularity in all of these pairs. From V_A^i we remove all exceptional vertices v_A for which

$$\deg_{G_1}(v_A, V_B^i) < \left(\frac{1}{r} - \varepsilon\right) m,$$

and from V_B^i all exceptional vertices v_B for which

$$\deg_{G_1}(v_B, V_A^i) < \left(\frac{1}{r} - \varepsilon\right) m.$$

(ε, G_1)-regularity guarantees that at most εm vertices are removed from each cluster. By doing this we may create some discrepancies in the cardinalities of the clusters of this connected matching. We remove some more vertices from clusters V_A^i and V_B^i to assure that now we have the same number of vertices left in each cluster of the matching. For simplicity we still keep the notation $f(e_i) = (V_A^i, V_B^i)$ for the modified clusters. The removed vertices are added to the exceptional set V_B^0.

To get the final cycle in G_1 will use the following property of (ε, δ, G)-super-regular pairs.

Lemma 2.8. *For every $\delta > 0$ there exist an $\varepsilon > 0$ and m_0 such that the following holds. Let G be a bipartite graph with bipartition $V(G) = V_1 \cup V_2$ such that $|V_1| = |V_2| = m \geq m_0$, and let the pair (V_1, V_2) be (ε, δ, G)-super-regular. Then for every pair of vertices $v_1 \in V_1, v_2 \in V_2$, G contains a Hamiltonian path connecting v_1 and v_2.*

A lemma somewhat similar to Lemma 2.8 is used by Łuczak in [Łuc99] and by Haxell in [Hax97]. Lemma 2.8 is a special case of the much stronger Blow-up Lemma (see [KSSz97] and [KSSz98]).

Applying Lemma 2.8 for $1 \leq i \leq l_1$, we get a path in $G_1|_{f(e_i)}$ connecting v_A^i and v_B^i that contains all of the remaining vertices of $f(e_i)$ (in case of $i = 1$ we just select a Hamiltonian path of $f(e_1)$ starting from v_B^1 and in case of $i = l_1$, we select a Hamiltonian path of $f(e_{l_1})$ starting from $v_A^{l_1}$). These paths together with the connecting paths give us the desired G_1 cycle.

We repeat this procedure for all the at most r monochromatic connected matchings. This gives us a covering of B with at most r vertex disjoint monochromatic cycles that cover B apart from at most $2\sqrt{\varepsilon}|B|$ vertices. For the covering of these remaining vertices we can apply the following lemma from [GRSSz06] (Lemma 8 in [GRSSz06]).

Lemma 2.9. *There exists a constant n_0 such that the following is true. Assume that the edges of the complete bipartite graph $K(A, B)$ are colored with r colors. If $|A| \geq n_0$, $|B| \leq |A|/(8r)^{8(r+1)}$, then B can be covered by at most $2r$ vertex disjoint monochromatic cycles.*

Indeed we can apply this lemma as ε is sufficiently small. Thus altogether we covered B with at most $r + 2r = 3r$ vertex disjoint monochromatic cycles, and thus finishing the proof of Theorem 1.2. □

2.3 Covering B by Monochromatic Connected Subgraphs

We show here that a covering of B with vertex disjoint connected monochromatic subgraphs is possible if $|B|$ is not too large compared to $|A|$. To achieve that, we need a generalization of the following result.

Theorem 2.10 (Haxell, Kohayakawa [HK96]). *Let $r \geq 1$ and $n \geq 3r^4 r!(1-1/r)^{3(1-r)} \log r$ be integers, and suppose the edges of K_n are colored with r colors. Then K_n contains $t \leq r$ monochromatic trees $T_1, \ldots, T_t$ of radius at most 2, each of different color, such that their vertex sets $V(T_i)$ $(1 \leq i \leq t)$ partition the vertex set of K_n.*

We shall prove that Theorem 2.10 remains true even if there is a not too large "hole" in K_n. More precisely, let $H = H(A,B)$ be the graph whose vertex set is partitioned into A and B and contains all edges except the ones inside B.

Theorem 2.11. *Let $r \geq 1$ and suppose the edges of $H = H(A,B)$ are colored with r colors, where $|A| = n$, $|B| < e^{n/5r^{r+3}} - n$ (in particular, n is sufficiently large). Then H contains $t \leq r$ vertex disjoint monochromatic trees $T_1, \ldots, T_t$ of radius at most 2, each of different color, such that their vertex sets $V(T_i)$ $(1 \leq i \leq t)$ partition the vertex set of H.*

Corollary 2.12. *Let $r \geq 1$ and suppose the edges of the complete bipartite graph $K(A,B)$ are colored with r colors, $|A| = n$. If $|B| < e^{n/5r^{r+3}} - n$ (in particular, n is sufficiently large) then B can be covered by the vertices of vertex disjoint monochromatic trees $\{T_1, \ldots, T_t\}$, $t \leq r$, of radius at most 2, each of different color.*

Proof. Consider an arbitrary coloring of the edges of $K(A,B)$ with r colors and color all $\binom{n}{2}$ edges inside A with a new color, say, $r+1$. This is an $(r+1)$-coloring of the edges of $H(A,B)$. Thus, by Theorem 2.11 it contains $t \leq r+1$ monochromatic trees $T_1, \ldots, T_t$ of radius at most 2, each of different color such that their vertex sets $V(T_i)$ $(1 \leq i \leq t)$ partition the vertex set of H. But color $r+1$ can be used only to cover some subset of vertices in A. Therefore the trees whose color is not $r+1$ have the required property. ◻

Proof of Theorem 2.11. We may assume $r \geq 2$, otherwise the statement is trivial. We tailor the proof of Haxell and Kohayakawa [HK96] to our needs. For some k, $1 \leq k \leq r$, a *k-anchor* is a k-edge colored complete bipartite graph $[X,Y]$ with $|X| = k$, $|Y| \geq s_k$ such that for $x_i \in X$ all edges of the form $[x_i, Y]$ are colored with color i $(i = 1, \ldots, k)$. Let $s_i = n/r^i$, for $1 \leq i \leq r$ and $s_i = 0$ for $i > r$. Clearly, the sequence s_i is non-increasing. Let $\Gamma_i(v,V)$ be the neighborhood of the vertex v in color i in some subset of vertices V, $d_i(v,V) = |\Gamma_i(v,V)|$.

Consider an arbitrary r-edge coloring of $H(A,B)$, $|A| = n$, and a t-anchor $[X,Y]$ such that $X \subseteq A \cup B$, $Y \subseteq A$ and maximal in the sense that no $(t+1)$-anchor $[X_1, Y_1]$ with $Y_1 \subseteq A$ exists in this coloring. Let $X = \{x_1, \ldots, x_t\}$

and assume $\{1,\dots,t\}$, are the colors of the t-anchor. Since a 1-anchor can be defined by selecting $x_1 \in B$ and defining color 1 as the majority color on $[x_1, A]$, t is well defined.

Now we proceed to prove that the vertices of $Z_0' = (A \cup B) \setminus (X \cup Y)$ can be covered by vertex disjoint monochromatic stars with centers in Y. In fact we achieve this by applying the following greedy procedure in less than

$$\lfloor s_r/2r \rfloor \le s_r/2r \le s_t/2r \le |Y| \tag{2}$$

steps.

Let $y_1 \in Y$ be the vertex which is adjacent to the most vertices in Z_0' in some color $i_1 \in [t]$ (i.e., we pick a monochromatic star centered in Y containing the most leaves in Z_0'). Let $Z_1 \subseteq Z_0'$ be the set of the leaves just chosen, $Z_1' = Z_0' \setminus Z_1$. In general, assume that vertices $y_1,\dots,y_q \in Y$, not necessarily different colors $i_1,\dots,i_q \in [t]$, pairwise disjoint sets $Z_1,\dots,Z_q$ and sets $Z_1',\dots,Z_q'$ are already defined. Let $Y_q = Y \setminus \{y_1,\dots,y_q\}$. Select $y_{q+1} \in Y_q$ and $i_{q+1} \in [t]$ such that $d_{i_{q+1}}(y_{q+1}, Z_q')$ is maximal, $Z_{q+1} = \Gamma_{i_{q+1}}(y_{q+1}, Z_q')$, $Z_{q+1}' = Z_q' \setminus Z_{q+1} = Z_0' \setminus (\cup_{i=1}^{q} Z_i)$.

Consider the edges between the (yet uncovered) vertices in Z_q' and the (yet not used) vertices in Y_q (Y_q is nonempty because of (2)). We have

$$\sum_{z\in Z_q'} \sum_{1\le i\le t} d_i(z, Y_q) > |Z_q'| \left(|Y| - q - (r-t)s_{t+1}\right).$$

Indeed, $|Y_q| = |Y| - q$, and a vertex $z \in Z_q'$ is adjacent to less than s_{t+1} vertices of Y in each color j, $t+1 \le j \le r$. Else, if $z \in Z_q'$, $Y^* \subset Y$, $|Y^*| \ge s_{t+1}$ exist such that all edges in $[z, Y^*]$ colored j, $t+1 \le j \le r$, then $\{z\} \cup X$ with Y^* would form a $(t+1)$-anchor, contradicting the choice of t. Therefore, by a standard averaging argument

$$\begin{aligned} d_{i_{q+1}}(y_{q+1}, Z_q') &\ge \frac{1}{t}\frac{1}{|Y|-q}|Z_q'| \left(|Y| - q - (r-t)s_{t+1}\right) \\ &= \frac{1}{t}|Z_q'| \left(1 - \frac{(r-t)s_{t+1}}{|Y|-q}\right). \end{aligned}$$

Using (2) we have

$$\begin{aligned} |Z_{q+1}'| = |Z_q'| - |Z_{q+1}| &\le |Z_q'| \left(1 - \frac{1}{t}\left(1 - \frac{(r-t)s_{t+1}}{|Y|-q}\right)\right) && (3) \\ &\le |Z_q'| \exp\left\{-\frac{1}{t}\left(1 - \frac{(r-t)s_{t+1}}{|Y|-q}\right)\right\} && (4) \\ &\le |Z_q'| \exp\left\{-\frac{1}{t}\left(1 - \frac{(r-1)s_{t+1}}{s_t - s_t/(2r)}\right)\right\} && (5) \\ &= |Z_q'| \exp\left\{-\frac{1}{(2r-1)t}\right\} \le |Z_q'| \exp\left\{-\frac{1}{2rt}\right\} && (6) \\ &\le |Z_q'| \exp\left\{-\frac{1}{2r^2}\right\}. && (7) \end{aligned}$$

To obtain (5) we utilized $|Y| \geq s_t$ and (2), and (6) follows from (5) by $s_t = rs_{t+1}$. Summarizing,

$$|Z'_{q+1}| \leq |Z'_q| \exp\left\{-\frac{1}{2r^2}\right\},$$

and we let our algorithm run for at most $\lfloor s_r/2r \rfloor$ steps. Therefore we shall cover all vertices in Z'_0 if

$$|Z'_0| \left(e^{-\frac{1}{2r^2}}\right)^{\left\lfloor \frac{s_r}{2r} \right\rfloor} \leq |Z'_0| \left(e^{-\frac{1}{2r^2}}\right)^{\left(\frac{s_r}{2r}-1\right)} = |Z'_0| \left(e^{-\frac{1}{2r^2}}\right)^{\left(\frac{n}{2r^{r+1}}-1\right)}$$
$$\leq |Z'_0| e^{-\frac{n}{5r^{r+3}}} < 1,$$

which is satisfied by

$$|Z'_0| < |V(H)| < e^{n/5r^{r+3}}. \tag{8}$$

Assume that we covered Z'_0 with monochromatic stars with centers $y_1, \dots, y_{q_0}$, colors $i_1, \dots, i_{q_0}$ and sets of leaves $Z_1, \dots, Z_{q_0}$. The partitioning trees $T_1, \dots, T_t$ of colors $1, \dots, t$ are defined as follows.

$$V(T_i) = \{x_i\} \cup \bigcup_{k \in [q_0]:\ i_k = i} (\{y_k\} \cup Z_k),$$

and

$$E(T_i) = \bigcup_{k \in [q_0]:\ i_k = i} (\{(x_i, y_k)\} \cup \{(y_k, z) : z \in Z_k\}).$$

Clearly, the vertices of $Y_{q_0} = Y \setminus \{y_1, \dots y_{q_0}\}$ can be added to, say, T_1. □

References

[AS91] N. Alon, J. H. Spencer, *The Probabilistic Method*, John Wiley and Sons, 1991.

[CL86] G. Chartrand, L. Lesniak, *Graph Theory*, Wadsworth, 1986.

[EGP91] P. Erdős, A. Gyárfás, L. Pyber, Vertex coverings by monochromatic cycles and trees, *Journal of Comb. Theory B* **51** (1991), 90–95.

[Für81] Z. Füredi, Maximum degree and fractional matchings in uniform hypergraphs, *Combinatorica* **1** (1981), pp. 155–162.

[Gya77] A. Gyárfás, *Partition Covers and Blocking sets in hypergraphs (in Hungarian)*, Thesis, Studies of Computer and Automation Research Institute of Hungarian Academy of Sciences, 177, (1977) MR 58, 5392.

[GRSSz06] A. Gyárfás, M. Ruszinkó, G. N. Sárközy, E. Szemerédi, Three-color Ramsey numbers for paths, to appear in *Combinatorica.*

[GRSSz06] A. Gyárfás, M. Ruszinkó, G. N. Sárközy, E. Szemerédi, An improved bound for the monochromatic cycle partition number, submitted. See also Worcester Polytechnic Institute, Computer Science Department, Technical Report 05-16:
`ftp://ftp.cs.wpi.edu/pub/techreports/pdf/05-16.pdf`

[Hax97] P. Haxell, Partitioning complete bipartite graphs by monochromatic cycles, *Journal of Comb. Theory B* **69** (1997), 210–218.

[HK96] P. Haxell, Y. Kohayakawa, Partitioning by monochromatic trees, *Journal of Comb. Theory B* **68** (1996), 218–222.

[KSSz97] J. Komlós, G. N. Sárközy, E. Szemerédi, Blow-up Lemma, *Combinatorica* **17** (1997), 109–123.

[KSSz98] J. Komlós, G. N. Sárközy, E. Szemerédi, An algorithmic version of the Blow-up Lemma, *Random Structures and Algorithms* **12** (1998), 297–312.

[KS96] J. Komlós, M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory, in D. Miklós, V. T. Sós, T. Szőnyi (eds) *Combinatorics, Paul Erdős is Eighty*, Bolyai Society Math. Studies, Vol.2, 295–352, Budapest, 1996.

[Łuc99] T. Łuczak, $R(C_n, C_n, C_n) \leq (4+o(1))n$, *Journal of Comb. Theory B* **75** (1999), 174–187.

Nonconstant Monochromatic Solutions to Systems of Linear Equations

Neil Hindman[1⋆] and Imre Leader[2]

[1] Department of Mathematics, Howard University, Washington, DC 20059, USA
nhindman@aol.com

[2] Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge, UK
I.Leader@dpmms.cam.ac.uk

Summary. The systems of linear equations (homogeneous or inhomogeneous) that are partition regular, over $\mathbb{N}$ or $\mathbb{Z}$ or $\mathbb{Q}$, were characterized by Rado. Our aim here is to characterize those systems for which we can guarantee a nonconstant, or injective, solution. It turns out that we thereby recover an equivalence between $\mathbb{N}$ and $\mathbb{Z}$ that is normally lost when one passes from homogeneous to inhomogeneous systems.

AMS Subject Classification. 05D10.
Keywords. Partition regular equations, Rado's theorem.

1 Introduction

We say that a $u \times v$ matrix A, with entries from $\mathbb{Q}$, is *partition regular* (or *kernel partition regular*) if whenever the positive integers $\mathbb{N}$ are finitely colored there is a vector $x \in \mathbb{N}^v$ that is monochromatic (meaning that all its entries are from the same color class) such that $Ax = 0$. We may also speak of the 'system of equations $Ax = 0$' being partition regular. Many of the classical results of Ramsey Theory may be interpreted as statements that particular matrices are partition regular. For example, Schur's Theorem [Sch16], that whenever $\mathbb{N}$ is finitely colored there exist x, y, z of the same color with $x + y = z$, is precisely the assertion that the 1×3 matrix $\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$ is partition regular.

The partition regular matrices were characterized by Rado in the 1930s [Rado33]. To give the characterization, we need to introduce another definition. Let the matrix A have columns $c_1, c_2, \ldots, c_v$. Then we say that A has the *columns property* if there is a partition of $\{1, 2, \ldots, v\}$ as $I_1 \cup I_2 \cup \ldots \cup I_m$ (some $m \geq 1$) such that

⋆ This author acknowledges support received from the National Science Foundation (USA) via grant DMS-0243586.

(1) $\sum_{i\in I_1} c_i = 0$; and
(2) for each $t > 1$, $\sum_{i\in I_t} c_i$ is a (rational) linear combination of $\{c_i : i \in I_1 \cup \ldots \cup I_{t-1}\}$.

Note that the columns property can be checked in finite time. Rado showed that a matrix is partition regular if and only if it has the columns property. (Although this paper is self-contained, the reader who wishes for background information may see [Deu89] or [GRS90].)

What happens over different spaces? We say that A is *partition regular over* $\mathbb{Z}$ (respectively $\mathbb{Q}$) if whenever $\mathbb{Z} \setminus \{0\}$ (respectively $\mathbb{Q} \setminus \{0\}$) is finitely colored there is a monochromatic vector x with $Ax = 0$. Then trivially if A is partition regular over $\mathbb{N}$, it is partition regular over $\mathbb{Z}$, and in fact the converse holds as well: indeed, if we have a bad k-coloring of $\mathbb{N}$ (meaning a coloring with k colors such that there is no monochromatic x with $Ax = 0$), then we may extend this to $\mathbb{Z}$ by coloring $-\mathbb{N}$ the same way as $\mathbb{N}$, but with k new colors – it is easy to check that this is a bad $2k$-coloring of $\mathbb{Z}$. It also turns out that partition regularity over $\mathbb{Z}$ and $\mathbb{Q}$ coincide, by a simple compactness argument (see [Deu89] for details).

Rado went on to consider *inhomogeneous* linear equations. Let A be a $u \times v$ matrix, and let $b \in \mathbb{Q}^u$. Then we say that the system of equations $Ax = b$ is *partition regular over* S (where S is one of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$) if whenever $S \setminus \{0\}$ is finitely colored there is a monochromatic vector x with $Ax = b$. In this inhomogenous setup, partition regularity over $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are definitely not the same. For example, the system $x + y + z = -6$ is partition regular over $\mathbb{Z}$ (just take $x = y = z = -2$) but not partition regular over $\mathbb{N}$.

Rado's characterization of partition regularity in the inhomogeneous case is as follows. If S is $\mathbb{Z}$ or $\mathbb{Q}$, then the system $Ax = b$ (with $b \neq 0$) is partition regular over S if and only if there is a constant solution. More precisely, writing $\overline{c}$ for the vector of the appropriate size all of whose coordinates are c, $Ax = b$ (with $b \neq 0$) is partition regular over S if and only if there exists $d \in S \setminus \{0\}$ such that $A\overline{d} = b$. In a sense, this is saying that if an inhomogeneous system is partition regular over $\mathbb{Z}$ or $\mathbb{Q}$ then it is partition regular for a trivial reason. Rado also showed that, over $\mathbb{N}$, the situation is 'halfway in between': the system $Ax = b$ ($b \neq 0$) is partition regular if and only if *either* there is a $d \in \mathbb{N}$ with $A\overline{d} = b$ *or* A has the columns property and there is a $d \in \mathbb{Z}$ with $A\overline{d} = b$. (These results are in [Rado33] and [Rado43]. To be precise, the cases of $\mathbb{N}$ and $\mathbb{Z}$ are in [Rado33], while the case of $\mathbb{Q}$, although not appearing explicitly, may easily be obtained from results in [Rado43].)

Our main aim in this paper is to consider what happens when we restrict our attention to *nonconstant* solutions ($Ax = b$ with x not a constant vector). There are two natural reasons for wanting to consider this question. Our first reason is that in fact some statements only appear artificially as partition regularity statements. Consider for example van der Waerden's Theorem [Wae27], which says that whenever $\mathbb{N}$ is finitely colored there exist arbitrarily long monochromatic arithmetic progressions. A natural statement of the length 5 instance of this theorem is that the equations

$$\begin{aligned} x_3 - x_2 &= x_2 - x_1 \\ x_4 - x_3 &= x_3 - x_2 \\ x_5 - x_4 &= x_4 - x_3 \end{aligned}$$

have a monochromatic solution *which is not constant*. The matrix corresponding to this system of equations is

$$\begin{pmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{pmatrix}$$

which satisfies the columns condition with $m = 1$, so Rado's Theorem only guarantees a constant solution to this system. (It can be made to guarantee a nonconstant solution by adding the equation $x_5 - x_4 = x_6$, that is by requiring that the increment also be the same color as the terms of the progression.) So, apart from the strengthening to insist that the increment is the same color, the natural way to have van der Waerden's Theorem as a partition regularity statement would be to introduce 'nonconstant' as an extra condition.

Our second reason concerns the inhomogeneous results over $\mathbb{Z}$ and $\mathbb{Q}$. The fact that the only way for a system $Ax = b$ to be partition regular is for there to be trivial (constant) solutions suggests that one is not asking the right question. Removing the constant solutions stops this particular phenomenon (and, as we shall see, gives a much richer structure to the characterization).

One rather unexpected consequence of restricting to nonconstant solutions is that it turns out that, even in the inhomogeneous case, partition regularity over $\mathbb{N}$ and $\mathbb{Z}$ now coincide. This will follow from our characterizations. Curiously, this rather pleasant feature seems not to have a direct proof. It seems remarkable that the equivalence could not be 'trivially obvious', given that it is true, but we have been unable to find a direct argument.

Instead of asking for nonconstant monochromatic solutions, one could also ask whether there are injective monochromatic solutions. (In a sense, this is what one really wants in the case of van der Waerden's Theorem. However, any nonconstant solution of those equations is automatically injective.) More generally, one can ask that certain specific coordinates in a monochromatic solution be distinct. We give characterizations here as well.

The plan of the paper is as follows. In Section 2 we present some preliminary results, including a proof of the case $S = \mathbb{Q}$ of the result of Rado on partition regularity of $Ax = b$ over $\mathbb{Q}$ – we include this for completeness, and to make the paper more readable. Then Section 3 contains our main results characterizing the existence of monchromatic nonconstant or injective solutions to $Ax = 0$ and $Ax = b$ in $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$.

2 Preliminaries

We begin by presenting a proof of the characterization of partition regularity of the system $Ax = b$ (where $b \neq 0$) over $\mathbb{Q}$. The proof is based on [GRS90, Lemma 22 and Corollary 24, pp. 87-88] and [Rado43, Lemma 4].

Lemma 2.1. *Let $v \in \mathbb{N}$. There is a coloring χ of $\mathbb{R}$ in $2v$ colors such that there do not exist $\langle x_j\rangle_{j=1}^v$ and $\langle y_j\rangle_{j=1}^v$ in $\mathbb{R}$ with $\chi(x_j) = \chi(y_j)$ for each $j \in \{1,2,\ldots,v\}$ and $\sum_{j=1}^v (x_j - y_j) = 1$.*

Proof. Define $\chi : \mathbb{R} \to \{0,1,\ldots,2v-1\}$ by $\chi(x) = i$ if and only if there is some $m \in \mathbb{Z}$ such that $2m + \dfrac{i}{v} \leq x < 2m + \dfrac{i+1}{v}$. Suppose one has $\langle x_j\rangle_{j=1}^v$ and $\langle y_j\rangle_{j=1}^v$ in $\mathbb{R}$ with $\chi(x_j) = \chi(y_j)$ for each $j \in \{1,2,\ldots,v\}$ and $\sum_{j=1}^v (x_j - y_j) = 1$. Then given j one has some $m_j \in \mathbb{Z}$ such that $2m_j - \frac{1}{v} < x_j - y_j < 2m_j + \frac{1}{v}$. Let $n = \sum_{j=1}^v m_j$. Then $2n-1 < 1 < 2n+1$, a contradiction. □

Lemma 2.2. *Assume that the equation $\sum_{j=1}^v c_j x_j = b$ is partition regular over $\mathbb{Q}$ where $b \in \mathbb{Q}$ and each $c_j \in \mathbb{Q}$. Then there is some $d \in \mathbb{Q}$ such that $d\sum_{j=1}^v c_j = b$. In particular, if $\sum_{j=1}^v c_j = 0$, then $b = 0$.*

Proof. If $\sum_{j=1}^v c_j \neq 0$, let $d = \dfrac{b}{\sum_{j=1}^v c_j}$. So assume that $\sum_{j=1}^v c_j = 0$ and suppose that $b \neq 0$. Define a coloring χ^* of $\mathbb{Q}$ by $\chi^*(x) = \chi^*(y)$ if and only if for each $j \in \{1,2,\ldots,v\}$, $\chi(\frac{c_j x}{b}) = \chi(\frac{c_j y}{b})$, where χ is as guaranteed by Lemma 2.1 for $v-1$. Pick monochrome $x_1, x_2, \ldots, x_v$ such that $\sum_{j=1}^v c_j x_j = b$. Then $\sum_{j=2}^v \left(\dfrac{c_j x_j}{b} - \dfrac{c_j x_1}{b}\right) = 1$, contradicting Lemma 2.1. □

Lemma 2.3. *Assume that A is a $2 \times v$ matrix with entries from $\mathbb{Q}$, $b \in \mathbb{Q}^2$, and the equation $Ax = b$ is partition regular over $\mathbb{Q}$. Then for any choice of $t_1, t_2 \in \mathbb{Q}$, there is some $d \in \mathbb{Q}$ such that $d(t_1 \sum_{j=1}^v a_{1,j} + t_2 \sum_{j=1}^v a_{2,j}) = t_1 b_1 + t_2 b_2$.*

Proof. Let $c_j = t_1 a_{1,j} + t_2 a_{2,j}$ and let $b = t_1 b_1 + t_2 b_2$. We claim that the equation $\sum_{j=1}^v c_j x_j = b$ is partition regular over $\mathbb{Q}$. So let $\mathbb{Q}$ be finitely colored and pick monochrome x such that $Ax = b$. Then $\sum_{j=1}^v c_j x_j = t_1 \sum_{j=1}^v a_{1,j} x_j + t_2 \sum_{j=1}^v a_{2,j} x_j = t_1 b_1 + t_2 b_2 = b$. Pick d as guaranteed by Lemma 2.2. □

Lemma 2.4. *Assume that A is a $2 \times v$ matrix with entries from $\mathbb{Q}$, $b \in \mathbb{Q}^2$, the equation $Ax = b$ is partition regular over $\mathbb{Q}$, $s_1 = \sum_{j=1}^v a_{1,j} \neq 0$, and $s_2 = \sum_{j=1}^v a_{2,j} \neq 0$. Then $\dfrac{b_1}{s_1} = \dfrac{b_2}{s_2}$.*

Proof. By Lemma 2.3, if $s_1t_1 + s_2t_2 = 0$, then $b_1t_1 + b_2t_2 = 0$ so the system

$$\begin{aligned} s_1t_1 + s_2t_2 &= 0 \\ b_1t_1 + b_2t_2 &= 1 \end{aligned}$$

is not solvable so $\begin{vmatrix} s_1 & s_2 \\ b_1 & b_2 \end{vmatrix} = 0$. Thus $\frac{b_1}{s_1} = \frac{b_2}{s_2}$ as required. □

Theorem 2.5 (Rado). *Let $u, v \in \mathbb{N}$, let A be a $u \times v$ matrix with entries from $\mathbb{Q}$, let $b \in \mathbb{Q}^u \setminus \{0\}$. The system $Ax = b$ is partition regular over $\mathbb{Q}$ if and only if there exists $d \in \mathbb{Q} \setminus \{0\}$ such that $A\overline{d} = b$.*

Proof. Given any $i \in \{1, 2, \ldots, u\}$ if $\sum_{j=1}^{v} a_{i,j} = 0$, then by Lemma 2.2, $b_i = 0$ so any choice of d will work for that row. If for all $i \in \{1, 2, \ldots, u\}$, $\sum_{j=1}^{v} a_{i,j} = 0$, then we are done. So assume we have some $i \in \{1, 2, \ldots, u\}$ such that $\sum_{j=1}^{v} a_{i,j} \neq 0$ and let $d = \frac{b_i}{\sum_{j=1}^{v} a_{i,j}}$. By Lemma 2.4, $A\overline{d} = b$. □

We shall use the notion of a *first entries matrix*, a notion based on the *mpc-sets* introduced by Deuber in [Deu73]. Let $u, v \in \mathbb{N}$ and let A be a $u \times v$ matrix with entries from $\mathbb{Q}$. Then A is a *first entries matrix* if and only if no row of A is 0 and whenever $i, j \in \{1, 2, \ldots, u\}$ and $k = \min\{t \in \{1,2,\ldots,v\} : a_{i,t} \neq 0\} = \min\{t \in \{1, 2, \ldots, v\} : a_{j,t} \neq 0\}$, then $a_{i,k} = a_{j,k} > 0$.

Lemma 2.6 (Deuber). *Let $u, v \in \mathbb{N}$ and let A be a $u \times v$ first entries matrix. Whenever $\mathbb{N}$ is finitely colored, there exists $x \in \mathbb{N}^v$ such that all entries of Ax are monochrome.*

Proof. This is essentially in [Deu73]. The proof may also be found in [HS98, Theorem 15.24]. □

A matrix satisfying the conclusion of Lemma 2.6 is said to be *image partition regular.*

Lemma 2.7. *Let $u, v \in \mathbb{N}$ and let A be a $u \times v$ first entries matrix. Whenever $\mathbb{N}$ is finitely colored, there exists $x \in \mathbb{N}^v$ such that all entries of Ax are monochrome and entries of Ax corresponding to unequal rows of A are distinct.*

Proof. By Lemma 2.6 A is image partition regular, so by [HLS02, Theorem 2.10] the conclusion holds. (Statement (n) of [HLS02, Theorem 2.10] refers to finding the image in a given *central* set. One only needs to know that given any finite partition of $\mathbb{N}$, one cell must be central.) □

3 Nonconstant Monochromatic Solutions

In this section we determine precisely those systems of homogeneous and those systems of inhomogeneous solutions which always have nonconstant or injective solutions whenever $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{Q}$ are finitely colored. Our characterizations typically state that a condition that is clearly necessary is in fact also sufficient. For example, for a matrix A to be nonconstant partition regular (over $\mathbb{N}$, $\mathbb{Z}$ or $\mathbb{Q}$) we certainly require that A has the columns property and also that there is some nonconstant linear dependence among the columns of A, and statement (e) of Theorem 3.2 asserts that this condition is also sufficient.

A key idea in the proofs will be the general Ramsey philosophy of 'if something can be forced, then it can be forced in a monochromatic way'. Thus for example if we wish to find solutions in which two particular variables x and y are distinct, we do not find such solutions directly, but rather we introduce a new variable z and a new equation $x + z = y$, so that in any solution of the new system we must have $x \neq y$.

Theorem 3.1. *Let $u, v \in \mathbb{N}$, let A be a $u \times v$ matrix with entries from $\mathbb{Q}$, and let $F \subseteq \{1, 2, \ldots, v\}$. The following statements are equivalent.*

(a) *Whenever $\mathbb{N}$ is finitely colored there exists monochromatic $x \in \mathbb{N}^v$ such that $Ax = 0$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*
(b) *Whenever $\mathbb{Z} \setminus \{0\}$ is finitely colored there exists monochromatic $x \in \mathbb{Z}^v$ such that $Ax = 0$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*
(c) *Whenever $\mathbb{Q} \setminus \{0\}$ is finitely colored there exists monochromatic $x \in \mathbb{Q}^v$ such that $Ax = 0$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*
(d) *The matrix A satisfies the columns condition and there exists $x \in \mathbb{Q}^v$ such that $Ax = 0$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*

Proof. That (a) implies (b) and (b) implies (c) is trivial. That (c) implies (d) follows immediately from Rado's Theorem.

To see that (d) implies (a) pick m, $\langle I_j \rangle_{j=1}^m$ and $\langle \delta_{i,t} \rangle_{i \in J_t}$ for $t \in \{2, 3, \ldots, m\}$ such that (as guaranteed by the columns condition) we have

(1) $\{I_1, I_2, \ldots, I_m\}$ is a partition of $\{1, 2, \ldots, v\}$;
(2) $\sum_{i \in I_1} c_i = 0$; and
(3) if $m > 1$ and $t \in \{2, 3, \ldots, m\}$, then $\sum_{i \in I_t} c_i = \sum_{i \in J_t} \delta_{i,t} \cdot c_i$, where $J_t = \bigcup_{j=1}^{t-1} I_j$.

Define a $v \times m$ matrix B by, for $i \in \{1, 2, \ldots, v\}$ and $j \in \{1, 2, \ldots, m\}$,

$$b_{i,j} = \begin{cases} 1 & \text{if } i \in I_j \\ -\delta_{i,j} & \text{if } i \in J_j \\ 0 & \text{otherwise.} \end{cases}$$

Then B is a first entries matrix and $AB = 0$. Pick $y \in \mathbb{Q}^v$ such that $Ay = 0$ and $y_i \neq y_j$ whenever i and j are distinct members of F. Let C be the $v \times (m+1)$ matrix whose first m columns are the columns of B and whose final column is y. Then C is a first entries matrix and the rows of C corresponding to members of F are distinct. Let $\mathbb{N}$ be finitely colored and pick by Lemma 2.7 some $x \in \mathbb{N}^{m+1}$ such that all entries of Cx are monochrome and entries of Cx corresponding to unequal rows of C are distinct. Let $z = Cx$. Then $Az = ACx = Ox = 0$. □

Notice that statement (d) of Theorem 3.1 is a computable condition. Notice also that by taking $F = \{1, 2, \ldots, v\}$ in Theorem 3.1 one has a characterization of matrices that are injectively kernel partition regular over $\mathbb{N}$, $\mathbb{Z}$, or $\mathbb{Q}$.

Observe that the single equation $a_1x_1 + a_2x_2 = 0$ is partition regular if and only if $a_1 = -a_2$, in which case there are no nonconstant solutions (unless $a_1 = a_2 = 0$). On the other hand Theorem 3.1 tells us that if $n > 3$, $a_1, a_2, \ldots, a_n \in \mathbb{Q} \setminus \{0\}$, and the equation $a_1x_1 + a_2x_2 + \ldots + a_nx_n = 0$ is partition regular, then it is injectively partition regular.

We have the following characterization of nonconstant partition regularity which includes a second (much more easily) computable condition.

Theorem 3.2. *Let $u, v \in \mathbb{N}$ and let A be a $u \times v$ matrix with entries from $\mathbb{Q}$. The following statements are equivalent.*

(a) *Whenever $\mathbb{N}$ is finitely colored there exists monochromatic nonconstant $x \in \mathbb{N}^v$ such that $Ax = 0$.*
(b) *Whenever $\mathbb{Z}$ is finitely colored there exists monochromatic nonconstant $x \in \mathbb{Z}^v$ such that $Ax = 0$.*
(c) *Whenever $\mathbb{Q}$ is finitely colored there exists monochromatic nonconstant $x \in \mathbb{Q}^v$ such that $Ax = 0$.*
(d) *The matrix A satisfies the columns condition and there exists nonconstant $x \in \mathbb{Q}^v$ such that $Ax = 0$.*
(e) *The matrix A satisfies the columns condition and if the sum of the columns of A is 0, then there exists nonempty $D \subsetneq \{1, 2, \ldots, v\}$ and for each $j \in D$ there exists $\alpha_j \in \mathbb{Q} \setminus \{0\}$ such that $\sum_{j \in D} \alpha_j c_j = 0$, where c_j is column j of A.*

Proof. As in the proof of Theorem 3.1 we have that (a) implies (b), (b) implies (c), and (c) implies (d).

To see that (d) implies (e), assume that $\sum_{j=1}^{v} c_j = 0$ and pick nonconstant $x \in \mathbb{Q}^v$ such that $\sum_{j=1}^{v} x_j c_j = 0$. Then $\sum_{j=2}^{v} (x_j - x_1) c_j = 0$. Let $D = \{j \in \{2, 3, \ldots, v\} : x_j \neq x_1\}$ and for $j \in D$ let $\alpha_j = x_j - x_1$.

To see that (e) implies (a) let m, $\langle I_j \rangle_{j=1}^{m}$, and B be as in the proof that (d) implies (a) in Theorem 3.1. If $m > 1$, pick $i \in I_1$ and $t \in I_2$, note that rows i and t of B are unequal, and let $C = B$. If $m = 1$, then pick nonempty $D \subsetneq \{1, 2, \ldots, v\}$ and for each $j \in D$ pick $\alpha_j \in \mathbb{Q} \setminus \{0\}$ such that $\sum_{j \in D} \alpha_j c_j = 0$. Define $y \in \mathbb{Q}^v$ by

$$y_j = \begin{cases} \alpha_j & \text{if } j \in D \\ 0 & \text{if } j \in \{1,2,\ldots,v\} \setminus D \end{cases}$$

and let C be the single column of B followed by y. Given $i \in D$ and $t \in \{1,2,\ldots,v\} \setminus D$, rows i and t of C are unequal.

In either case C is a first entries matrix with two unequal rows such that $AC = 0$. Let $\mathbb{N}$ be finitely colored and pick by Lemma 2.7 some $x \in \mathbb{N}^{m+1}$ such that all entries of Cx are monochrome and entries of Cx corresponding to unequal rows of C are distinct. Let $z = Cx$. Then $Az = ACx = 0x = 0$. □

We now turn our attention to nonconstant monochromatic solutions to inhomogeneous systems of linear equations.

Theorem 3.3. *Let $u, v \in \mathbb{N}$, let A be a $u \times v$ matrix with entries from $\mathbb{Q}$, let $b \in \mathbb{Q}^u \setminus \{0\}$, and let $F \subseteq \{1,2,\ldots,v\}$ with $|F| \geq 2$. If $S = \mathbb{Z}$ or $S = \mathbb{Q}$, then the following statements are equivalent.*

(a) *Whenever $S \setminus \{0\}$ is finitely colored there exists monochromatic $x \in S^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*
(b) *There exists $d \in S \setminus \{0\}$ such that $A\overline{d} = b$, A satisfies the columns condition, and there exists $x \in \mathbb{Q}^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*

Proof. To see that (a) implies (b), note that we may pick $d \in S \setminus \{0\}$ such that $A\overline{d} = b$ by Theorem 2.5 and trivially $x \in \mathbb{Q}^v$ exists as required. So it suffices to show that the system $Ax = 0$ is partition regular over S. So let $r \in \mathbb{N}$ and let $\varphi : S \setminus \{0\} \to \{1,2,\ldots,r\}$. Define $\psi : S \setminus \{0\} \to \{1,2,\ldots,r+1\}$ by

$$\psi(x) = \begin{cases} \varphi(x-d) & \text{if } x \neq d \\ r+1 & \text{if } x = d\,. \end{cases}$$

Pick $x \in S^v$ such that x is monochromatic with respect to ψ, $Ax = b$, and $x_i \neq x_j$ whenever i and j are distinct members of F. Since $|F| \geq 2$, the constant value of $\psi(x_i)$ cannot be $r+1$. Let $y = x - \overline{d}$. Then y is monochromatic with respect to φ, $y_i \neq y_j$ whenever i and j are distinct members of F, and $Ay = Ax - A\overline{d} = 0$.

To see that (b) implies (a), pick $d \in S \setminus \{0\}$ such that $A\overline{d} = b$ and pick $x \in \mathbb{Q}^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F. Let $r \in \mathbb{N}$ and let $\varphi : S \setminus \{0\} \to \{1,2,\ldots,r\}$. Define $\psi : S \setminus \{0\} \to \{1,2,\ldots,r+1\}$ by

$$\psi(y) = \begin{cases} \varphi(y+d) & \text{if } y \neq -d \\ r+1 & \text{if } y = -d\,. \end{cases}$$

Now $x - \overline{d} \in \mathbb{Q}^v$ and $A(x - \overline{d}) = 0$ so by Theorem 3.1 we may pick $z \in S^v$ such that z is monochromatic with respect to ψ, $Az = 0$, and $z_i \neq z_j$ when i and j are distinct members of F. Since $|F| \geq 2$, the constant value of $\psi(z_i)$ cannot be $r+1$. Let $y = z + \overline{d}$. Then y is monochromatic with respect to φ, $y_i \neq y_j$ whenever i and j are distinct members of F, and $Ay = Az + A\overline{d} = b$. □

Notice that Theorem 3.3 tells us that the single equation $2x_1 - 2x_2 + 2x_3 = 1$ is nonconstantly partition regular over $\mathbb{Q}$ but not over $\mathbb{Z}$. On the other hand the next theorem tells us that nonconstant partition regularity over $\mathbb{Z}$ is equivalent to nonconstant partition regularity over $\mathbb{N}$. As we stated earlier, we have been unable to find a trivial proof of this equivalence.

Theorem 3.4. *Let $u, v \in \mathbb{N}$, let A be a $u \times v$ matrix with entries from $\mathbb{Q}$, let $b \in \mathbb{Q}^u \setminus \{0\}$, and let $F \subseteq \{1, 2, \ldots, v\}$ with $|F| \geq 2$. The following statements are equivalent.*

(a) *Whenever $\mathbb{N}$ is finitely colored there exists monochromatic $x \in \mathbb{N}^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*
(b) *Whenever $\mathbb{Z} \setminus \{0\}$ is finitely colored there exists monochromatic $x \in \mathbb{Z}^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*
(c) *There exists $d \in \mathbb{Z} \setminus \{0\}$ such that $A\overline{d} = b$, A satisfies the columns condition, and there exists $x \in \mathbb{Q}^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F.*

Proof. Trivially (a) implies (b) and (b) implies (c) by Theorem 3.3. To see that (c) implies (a) pick $d \in \mathbb{Z} \setminus \{0\}$ such that $A\overline{d} = b$ and pick $x \in \mathbb{Q}^v$ such that $Ax = b$ and $x_i \neq x_j$ whenever i and j are distinct members of F. Let $r \in \mathbb{N}$ and let $\varphi : \mathbb{N} \to \{1, 2, \ldots, r\}$. If $d > 0$, define $\psi : \mathbb{N} \to \{1, 2, \ldots, r\}$ by $\psi(y) = \varphi(y + d)$. If $d < 0$, define $\psi : \mathbb{N} \to \{1, 2, \ldots, r - d\}$ by

$$\psi(y) = \begin{cases} \varphi(y + d) & \text{if } y > -d \\ r + y & \text{if } y \leq -d\,. \end{cases}$$

Now $x - \overline{d} \in \mathbb{Q}^v$ and $A(x - \overline{d}) = 0$ so by Theorem 3.1 we may pick $z \in \mathbb{N}^v$ such that z is monochromatic with respect to ψ, $Az = 0$, and $z_i \neq z_j$ when i and j are distinct members of F. Since $|F| \geq 2$, the constant value of $\psi(z_i)$ cannot be $r + t$ for any $t \leq -d$. Let $y = z + \overline{d}$. Then y is monochromatic with respect to φ, $y_i \neq y_j$ whenever i and j are distinct members of F, and $Ay = Az + A\overline{d} = b$. □

We close by remarking that it would be very nice to find a direct short proof for the fact proved above that the notions of nonconstant partition regularity for $Ax = b$ over $\mathbb{N}$ and $\mathbb{Z}$ are the same.

References

[BDHL94] V. Bergelson, W. Deuber, N. Hindman, and H. Lefmann, Rado's Theorem for commutative rings, *J. Comb. Theory (Series A)* **66** (1994), 68–92.

[Deu73] W. Deuber, Partitionen und lineare Gleichungssysteme, *Math. Zeit.* **133** (1973), 109–123.

[Deu89] W. Deuber, Developments based on Rado's dissertation 'Studien zur Kombinatorik', in *Surveys in Combinatorics, 1989*, (Siemons, J., ed.), Cambridge University Press (1989), pp. 52–74.

[GRS90] R. Graham, B. Rothschild, and J. Spencer, *Ramsey Theory*, Wiley, New York, 1990.

[HLS02] N. Hindman, I. Leader, and D. Strauss, Image partition regular matrices – bounded solutions and preservation of largeness, *Discrete Math.* **242** (2002), 115–144.

[HS98] N. Hindman and D. Strauss, *Algebra in the Stone–Čech compactification: theory and applications*, de Gruyter, Berlin, 1998.

[Rado33] R. Rado, Studien zur Kombinatorik, *Math. Zeit.* **36** (1933), 242–280.

[Rado43] R. Rado, Note on combinatorial analysis, *Proc. London Math. Soc.* **48** (1943), 122–160.

[Sch16] I. Schur, Über die Kongruenz $x^m + y^m = z^m (\mathrm{mod}\, p)$, *Jahresbericht der Deutschen Math.-Verein.* **25** (1916), 114–117.

[Wae27] B. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw Arch. Wiskunde **19** (1927), 212–216.

On the Induced Ramsey Number $IR(P_3, H)$

Alexandr Kostochka[1⋆] and Naeem Sheikh[2]

[1] University of Illinois, Urbana, IL, 61801 USA, and Institute of Mathematics, Novosibirsk, 630090, Russia
kostochk@math.uiuc.edu
[2] University of Illinois, Urbana, IL, 61801 USA
nsheikh@math.uiuc.edu

Summary. The induced Ramsey number $IR(G, H)$ is the least positive integer N such that there exists an N-vertex graph F with the property that for each 2-coloring of its edges with red and blue, either some induced in F subgraph isomorphic to G has all its edges colored red, or some induced in F subgraph isomorphic to H has all its edges colored blue. In this paper, we study $IR(P_3, H)$ for various H, where P_3 is the path with 3 vertices. In particular, we answer a question by Gorgol and Łuczak by constructing a family $\{H_n\}_{n=1}^{\infty}$ such that $\lim_{n\to\infty} \sup \frac{IR(P_3,H_n)}{IR_w(P_3,H_n)} > 1$, where $IR_w(G, H)$ is defined almost as $IR(G, H)$, with the only difference that G should be induced only *in the red subgraph* of F (not in F itself) and H should be induced only in the blue subgraph of F.

AMS Subject Classification. 05D10, 05C35.
Keywords. Induced Ramsey number, weak induced Ramsey number.

1 Introduction

The *induced Ramsey number*, $IR(G, H)$, is the greatest positive integer N such that for each graph F on at most $N - 1$ vertices, there exists a 2-coloring of its edges with red and blue such that no induced copy of G in F has all its edges red and no induced copy of H in F has all its edges blue. Say that a graph F is an *IR-graph for graphs G and H*, if for each 2-coloring of edges of F with red and blue, either some induced in F subgraph isomorphic to G has all its edges colored red, or some induced in F subgraph isomorphic to H has all its edges colored blue. In these terms, the *induced Ramsey number*, $IR(G, H)$, is the least order of an IR-graph for G and H. The fact that an IR-graph exists for each G and H and thus $IR(G, H)$ is finite was proved independently

⋆ The research of the first author was partially supported by the NSF grant DMS-0400498 and grant 03-01-00796 of the Russian Foundation for Basic Research.

by Deuber [Deu75], Erdős et al. [EHP75], and Rödl [Rödl73]. Estimating induced Ramsey numbers of graphs in different classes attracted considerable attention (see, e.g., [Deu75, GŁ02, HNR83, KPR98, ŁR96, HKŁ95, Neš95]). In particular, Haxell, Kohayakawa and Łuczak [HKŁ95] showed that the diagonal induced Ramsey numbers for paths and cycles grow linearly in terms of their lengths. However, there are very few exact results.

A characteristic similar to the induced Ramsey number is the *weak induced Ramsey number*, $IR_w(G, H)$ – the least positive integer N such that there exists an N-vertex graph F with the property that for each 2-coloring of its edges with red and blue, either the red subgraph of F contains an induced (in this red graph) copy of G, or the blue subgraph of F contains an induced (in this blue graph) copy of H. Gorgol and Łuczak [GŁ02] gave an example of a pair of graphs for which the induced Ramsey number is greater than the weak induced Ramsey number. Namely, they showed that $IR(P_3, P_4) = 7$ and $IR_w(P_3, P_4) = 6$, where P_k is the path with k vertices. They also asked whether there exists a sequence $\{H_n\}_{n=1}^{\infty}$ of graphs such that for some graph G,

$$\lim_{n\to\infty} \sup \frac{IR(G, H_n)}{IR_w(G, H_n)} > 1. \tag{1}$$

Among other results, Gorgol and Łuczak proved that for every $n \geq 3$,

$$1.5n - 1 \leq IR(P_3, P_n) \leq 2n - 1 \quad \text{and} \quad 4n/3 \leq IR_w(P_3, P_n) \leq 5n/3.$$

In this paper, we estimate $IR(P_3, H)$ for various graphs H. We give the general bound

$$IR(P_3, H) \leq |V(H)| + |E(H)| \tag{2}$$

and show that this bound is sharp when H is the union of complete graphs. Then we refine bound (2) for graphs having vertices with equal neighborhoods and prove that this refined bound is sharp when H is any complete multipartite graph or the disjoint union of complete multipartite graphs. We also answer in the affirmative the above question of Gorgol and Łuczak by constructing a sequence $\{H_n\}_{n=1}^{\infty}$ of graphs such that (1) holds for them with P_3 in place of G.

The structure of the paper is as follows. In the next section we give upper bounds on $IR(P_3, H)$ and prove that for some graphs they are exact. In the last section we answer the question of Gorgol and Łuczak [GŁ02].

2 Upper Bounds on $IR(P_3, H)$

Since P_3 is a very simple graph, $IR(P_3, H)$ grows at most linearly with the growth of $|V(H)| + |E(H)|$. A simple construction below proves this.

Theorem 2.1. *For every graph H, $IR(P_3, H) \leq |V(H)| + |E(H)|$.*

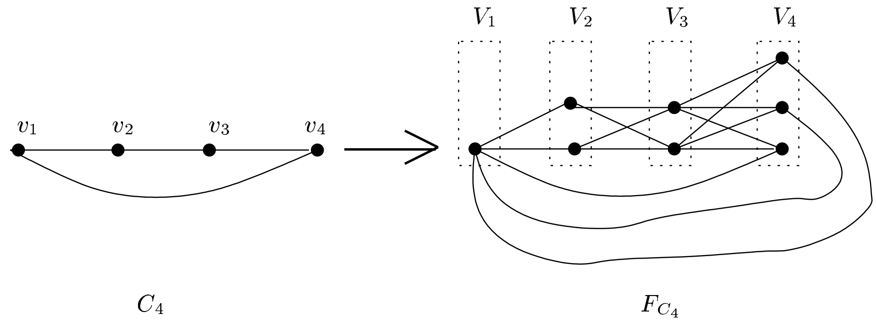


Fig. 1. An IR-graph for P_3 and C_4

Proof. Given a graph H, we construct the associated graph F_H as follows. Let $L = (v_1, v_2, \ldots, v_n)$ be a list of all vertices of H written in some order. For $i = 1, 2, \ldots, n$, let $d_{H,L}(v_i)$ be the number of neighbors of v_i in $\{v_1, \ldots, v_{i-1}\}$. The vertex set of $F = F_{H,L}$ is $V(F_{H,L}) = V_1 \cup V_2 \cup V_3 \cup \ldots \cup V_n$, where $|V_i| = 1 + d_{H,L}(v_i)$. For every edge (v_i, v_j) in H, we add all the edges between V_i and V_j in $F_{H,L}$. This completes the construction of F. Figure 1 illustrates this construction for C_4 in a particular list.

Note that $|V(F_{H,L})| = |V(H)| + \sum_{i=1}^{n} d_{H,L}(v_i) = |V(H)| + |E(H)|$.

Claim 2.2. Each red-blue edge coloring of $F_{H,L}$ contains either an induced red copy of P_3 or an induced blue copy of H such that $v_i \in V_i$ for every i.

Proof. We use induction on $n = |V(H)|$. The claim is trivially true for $n = 1$, in which case $H = F_{H,L} = K_1$. Suppose that the claim holds for each graph with less than n vertices. Consider a graph H with n vertices and let $L = (v_1, v_2, \ldots, v_n)$ be a list of all vertices of H. Let f be a red–blue edge coloring of $F_{H,L}$. Consider the graph $H' = H - v_n$ and let L' be the list of vertices of H' obtained from L by deleting v_n. Then $F_{H',L'} = F_{H,L} - V_n$. Let f' be the edge coloring induced in H' by f. Assume that $F_{H,L}$ has no induced red P_3. Then, as a subgraph of $F_{H,L}$, the graph $F_{H',L'}$ also has no induced red P_3. Thus, by the induction hypothesis, $F_{H',L'}$ has an induced blue copy of H' such that $v_i \in V_i$ for $i = 1, \ldots, n-1$. Let v_n have m neighbors in H. Then, $|V_n| = m+1$. Let M be the set of m vertices in the induced blue copy $\widetilde{H}$ of H' in $F_{H',L'}$ that need a new common neighbor to make the graph H. Each vertex in V_n is a potential candidate for this neighbor. If each of the $m + 1$ vertices in V_n has at least one red edge leading to M, then by pigeonhole principle, some vertex in M has two neighbors in V_n with the corresponding edges being red. Since V_n forms an independent set in $F_{H,L}$, this gives an induced red copy of P_3, a contradiction. Hence, at least one of the vertices in V_n, has all of its edges to M in blue color, thereby giving us an induced blue copy of H with $v_n \in V_n$. This proves the claim and thus the theorem. □

The following simple fact observed in [GŁ02] will be used for lower bounds on $IR(P_3, H)$.

Lemma 2.3. *Let F and H be any graphs and f be any red–blue edge coloring of F. If an edge $uv \in E(F)$ is colored red, then at most one of u and v can belong to an induced in F blue copy of H. As a consequence, any blue induced copy of H in F contains at most one vertex from each red clique in F.*

We now prove that the bound of Theorem 2.1 is sharp for the disjoint unions of complete graphs. The sign + between graphs below denotes the disjoint union of corresponding graphs.

Theorem 2.4. *For any positive integers $n_1 \leq \ldots \leq n_m$,*

$$IR(P_3, K_{n_1} + K_{n_2} + \ldots + K_{n_m}) = \sum_{i=1}^{m} \frac{n_i(n_i+1)}{2}.$$

Proof. Let $H = K_{n_1} + K_{n_2} + \ldots + K_{n_m}$. The upper bound follows from Theorem 2.1. Choose an IR-graph F for P_3 and H with fewest vertices.

We will make n_m attempts to color the edges of F. Let f_1 be the coloring of all the edges of F with blue. Since F is an IR-graph for P_3 and H, there is an induced copy H_1 of H. Recall that $|V(H_1)| = n_1 + n_2 + \ldots + n_m$. Color the edges of H_1 with red and all other edges with blue. This is f_2. Again, by the choice, F contains an induced copy H_2 of H. Let $H_{1,2} = H_1 - V(H_2)$. Since all the edges of H_1 are red, by Lemma 2.3, at most one vertex from each clique in H_1 belongs to $V(H_2)$. Hence $|V(H_{1,2})| \geq (n_1 - 1) + \ldots + (n_m - 1)$ and

$$|V(F)| \geq |V(H_{1,2})| + |V(H_2)| \geq \sum_{i=1}^{m} (n_i + (n_i - 1)).$$

Color the edges of H_2 and of $H_{1,2}$ with red and all other edges of F with blue. This is f_3. Again, by the choice, F contains an induced copy H_3 of H. And we do this way m times in total.

In general, after the kth attempt, we have a new blue induced in F copy H_k of H and $k-1$ partially destroyed copies of H: for $i = 1, \ldots, k-1$, let $H_{i,k} = H_i - V(H_{i+1}) - V(H_{i+2}) - \ldots - V(H_k)$. The subgraphs $H_{i,k}$ are vertex disjoint from each other and from H_k. Furthermore, by Lemma 2.3, for every $i = 1, \ldots, k-1$, each clique in $H_{i,k-1}$ has at most one vertex in common with H_k. Therefore,

$$|V(H_{i,k})| \geq \max\{0, n_1 - (k-i)\} + \ldots + \max\{0, n_m - (k-i)\}$$

and hence

$$|V(F)| \geq \sum_{i=1}^{k} \sum_{j=1}^{m} \max\{0, n_j - (k-i)\} = \sum_{j=1}^{m} \sum_{l=0}^{k-1} \max\{0, n_j - l\}. \qquad (3)$$

Thus, after the n_mth attempt, (3) yields

$$|V(F)| \geq \sum_{j=1}^{m} \sum_{l=0}^{n_m} \max\{0, n_j - l\} = \sum_{j=1}^{m} \sum_{l=0}^{n_j} (n_j - l) = \sum_{j=1}^{m} \frac{n_j(n_j+1)}{2}.$$

This proves the theorem. □

The bound might be sharp for some other graphs, but it is not sharp for graphs having vertices with the same nontrivial neighborhood, like complete multipartite graphs. For such graphs, we modify the bound.

Let H be a graph. Say that vertices v and w *are equivalent* (and write $v \sim w$) if their neighborhoods are the same. In particular, equivalent vertices are not adjacent. Partition $V(H)$ into the equivalence classes $V(H) = W_1 \cup \ldots W_s$ so that $|W_1| \leq \ldots \leq |W_s|$. Let $L_W = (v_1, v_2, \ldots, v_n)$ be a list of vertices of H such that it first encounters vertices in W_1, then in W_2, and so on. By the construction, all "degrees to the left" $d_{H,L_W}(v_i)$ are the same for vertices in the same equivalence class.

Theorem 2.5. *For every graph H, and any choice of $w_j \in W_j$ for $j = 1, \ldots, s$,*

$$IR(P_3, H) \leq |V(H)| + \sum_{j=1}^{s} d_{H,L_W}(w_j).$$

Proof. Given a graph H, we construct the graph $F' = F'_{H,L_W}$ as follows. The vertex set of F' is $V(F'_{H,L_W}) = V_1 \cup V_2 \cup V_3 \cup \ldots \cup V_s$, where $|V_j| = |W_j| + d_{H,L_W}(w_j)$. For every edge (w_i, w_j) in H, we add all the edges between V_i and V_j in F'_{H,L_W}. This construction is illustrated for the graph C_4 in Fig. 2 (compare with Fig. 1).

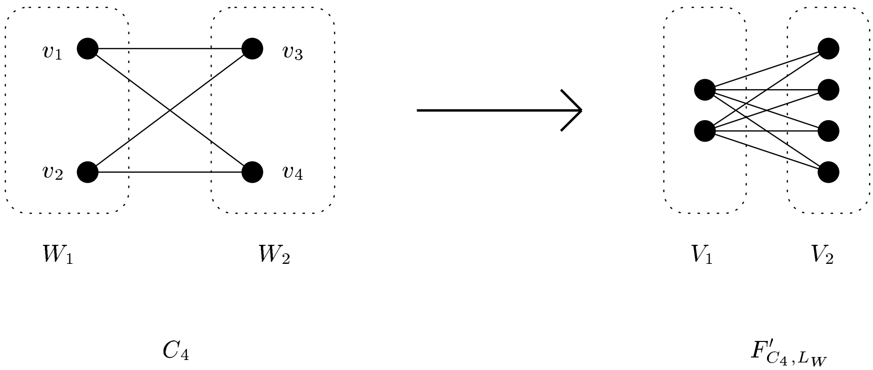


Fig. 2. Another IR-graph for P_3 and C_4

Note that $|V(F'_{H,L})| = |V(H)| + \sum_{j=1}^{s} d_{H,L_W}(w_j)$.

Claim 2.6. For each red-blue edge coloring of F'_{H,L_W}, it contains either an induced red copy of P_3 or an induced blue copy of H such that for every $j = 1, \ldots, s$, each $w_i \in W_j$ belongs to V_j.

Proof. The proof is by induction on s and is very similar to that of Claim 2.2. The claim is trivially true for $s = 1$, in which case H and F' are equal edgeless graphs. Suppose that the claim holds for each graph with less than s equivalence classes. Consider a graph H with s equivalence classes. Let f be a red–blue edge coloring of F'. Consider the graph $H' = H - W_s$ and let L'_W be the list of vertices of H' obtained from L_W by deleting W_s. Then $F'_{H',L'_W} = F'_{H,L_W} - W_s$. Assume that F'_{H,L_W} has no induced red P_3. Then, as a subgraph of F'_{H,L_W}, the graph F'_{H',L'_W} also has no induced red P_3. Thus, by the induction hypothesis, F'_{H',L'_W} has an induced blue copy of H' such that for every $j = 1, \ldots, s-1$ each $w_i \in W_j$ belongs to V_j. Let w_s have m neighbors in H. Then, $|V_s| = m + |W_s|$. Let M be the set of m vertices in the induced blue copy $\widetilde{H}$ of H' in F'_{H',L'_W} that need $|W_s|$ new common neighbors to make the graph H. Each of the $m + |W_s|$ vertices in V_n is a potential candidate for such a neighbor. If at least $m + 1$ of the $m + |W_s|$ vertices in V_s have at least one red edge leading to M, then, by pigeonhole principle, some vertex in M has two neighbors in V_s with the corresponding edges being red. Since V_s forms an independent set in F'_{H,L_W}, this gives an induced red copy of P_3, a contradiction. Hence, at least $|W_s|$ of the vertices in V_s, have all their edges to M in blue, thereby giving us an induced blue copy of H with every vertex of W_s in V_s. This proves the claim and thus the theorem. □

Remark. For graphs with large equivalence classes, the bound of Theorem 2.5 is significantly better than that of Theorem 2.1. For example, Theorem 2.1 yields $IR(P_3, K_{m,m}) \leq 2m + m^2$, while Theorem 2.5 gives a stronger bound of $IR(P_3, K_{m,m}) \leq 3m$. In fact, the bound of Theorem 2.5 is tight for complete multipartite graphs.

Theorem 2.7. *Let $n_1 \leq n_2 \leq \ldots \leq n_s$ be positive integers and $H = K_{n_1,n_2,\ldots,n_s}$. Then*

$$IR(P_3, H) = sn_1 + (s-1)n_2 + \ldots + n_s = \sum_{i=1}^{s} n_i(s+1-i).$$

Proof. The upper bound follows from Theorem 2.5. Choose an IR-graph F for P_3 and H with fewest vertices.

We will make s attempts to color the edges of F. Let f_1 be the coloring of all the edges of F with blue. Since F is an IR-graph for P_3 and H, there is an induced copy H_1 of H. Recall that $|V(H_1)| = n_1 + n_2 + \ldots + n_s$. Let H'_1 be a spanning subgraph of H_1 which is the disjoint union of n_1 cliques of size s, and $n_2 - n_1$ cliques of size $(s-1)$, and so on all the way down to $n_s - n_{s-1}$ cliques of size 1. Color the edges of H'_1 with red and all other edges with blue. This is f_2. Again, by the construction and the choice of F, it contains an induced copy H_2 of H. Let $H_{1,2} = H_1 - V(H_2)$. Since all the edges of H'_1 are red, by Lemma 2.3, the set $V(H'_1) \cap V(H_2)$ is independent in H'_1. Recall that H'_1 has n_s disjoint cliques. Hence $|V(H_{1,2})| \geq n_1 + \ldots + n_{s-1}$ and

$$|V(F)| \geq |V(H_{1,2})| + |V(H_2)| \geq n_s + 2\sum_{i=1}^{s-1} n_i.$$

Let H_2' be a subgraph of H_2 isomorphic to H_1'. Color the edges of H_2' and of $H_{1,2}$ with red and all other edges of F with blue. This is f_3. Again, F contains an induced copy H_3 of H. And we do this way s times in total.

In general, after the kth attempt, we have a new induced in F blue copy H_k of H and $k-1$ partially destroyed copies of H: for $i = 1, \ldots, k-1$, let $H_{i,k} = H_i' - V(H_{i+1}) - V(H_{i+2}) - \ldots - V(H_k)$, where H_i' is a spanning subgraph of H_i which is a disjoint union of n_s cliques isomorphic to H_1'. The subgraphs $H_{i,k}$ are vertex disjoint from each other and from H_k. Furthermore, by Lemma 2.3, for every $i = 1, \ldots, k-1$, $H_i' - V(H_{i,k})$ is the union of $k-i$ independent sets in H_i'. By the construction of H_1', such union can contain at most $n_s + n_{s-1} + \ldots + n_{s-k+i+1}$ vertices. Therefore, $|V(H_{i,k})| \geq n_1 + \ldots + n_{s-k+i}$ and hence

$$|V(F)| \geq \sum_{i=1}^{k} \sum_{j=1}^{s-k+i} n_j = n_s + 2n_{s-1} + \ldots + (k-1)n_{s-k+2} + k \sum_{i=1}^{s-k+1} n_i. \quad (4)$$

Thus, after the sth attempt, (4) yields

$$|V(F)| \geq n_s + 2n_{s-1} + \ldots + sn_1.$$

This proves the theorem. □

In fact, the bound of Theorem 2.5 is exact for all disjoint unions of multipartite graphs.

Theorem 2.8. *Let $n_{1,1} \leq n_{1,2} \leq \ldots \leq n_{1,s_1}$, $n_{2,1} \leq n_{2,2} \leq \ldots \leq n_{2,s_2}, \ldots,$ $n_{m,1} \leq n_{m,2} \leq \ldots \leq n_{m,s_m}$ be positive integers. Let H be the disjoint union of the complete multipartite graphs $H_1 = K_{n_{1,1},n_{1,2},\ldots,n_{1,s_1}}$, $H_2 = K_{n_{2,1},n_{2,2},\ldots,n_{2,s_2}}, \ldots, H_m = K_{n_{m,1},n_{m,2},\ldots,n_{m,s_m}}$. Then*

$$IR(P_3, H) = \sum_{i=1}^{m} IR(P_3, H_i).$$

The upper bound immediately follows from Theorem 2.5 and the proof of the lower bound practically repeats that of Theorem 2.7 only with more subscripts, so we leave it to the reader.

3 Weak Versus Ordinary

As it was mentioned in the introduction, Gorgol and Łuczak [GŁ02] proved that $IR_w(P_3, P_4) = 6 < IR(P_3, P_4) = 7$. To prove the upper bound on $IR_w(P_3, P_4)$, they made the following observation.

Claim 3.1. For each red-blue coloring of the edges of the graph F_1 in Fig. 3 such that the red subgraph has no induced P_3, the blue subgraph has an induced path connecting the vertices of degree two. In particular, it contains an induced in blue subgraph P_4 starting at any vertex of degree two.

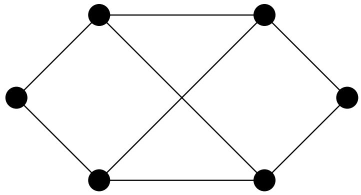


Fig. 3. The Gorgol–Łuczak example

Recall a couple of definitions. Let F be a graph. For a set $T \subset V(F)$, let $o(F-T)$ be the number of *odd* components of $F-T$, i.e. components of odd order, and let $\mathrm{def}(T) = o(F-T) - |T|$ be called *the deficiency of* T. The *deficiency* of F is

$$\mathrm{def}(F) = \max_{W \subseteq V(F)} \mathrm{def}(W) = \max_{W \subseteq V(F)} \{o(F-W) - |W|\}.$$

Let $\pi(F)$ denote the size of a maximum matching in F. By Berge–Tutte Formula, $\mathrm{def}(F) = |V(F)| - 2\pi(F)$.

The main result of this section confirming that (1) holds is the following.

Theorem 3.2. *For a positive integer k, let H_k be the vertex disjoint union of k paths P_4. Then $IR_w(P_3, H_k) \leq 6k$ and $IR(P_3, H_k) \geq 6.1k$. In particular, $\frac{IR(P_3,H_k)}{IR_w(P_3,H_k)} \geq 1 + 1/60$ for each positive integer k.*

Proof. The upper bound on $IR_w(P_3, H_k)$ is easy: we let F_k be the disjoint union of k copies of the 6-vertex graph in Fig. 3. As observed by Gorgol and Łuczak, for any red-blue coloring of the edges of F_k, each copy either contains induced in red P_3 or induced in blue P_4. Thus, the whole F_k either contains induced in red P_3 or induced in blue H_k. The lower bound on $IR(P_3, H_k)$ needs more work.

For a contradiction, suppose that $IR(P_3, H_k) = (6+\varepsilon)k$, where $\varepsilon < 0.1$ (and possibly is negative). By definition, εk is an integer. Consider a graph F with $N = (6+\varepsilon)k$ vertices such that for each red-blue coloring of its edges, either some induced in F subgraph isomorphic to P_3 has all its edges colored red, or some induced in F subgraph isomorphic to H_k has all its edges colored blue.

Lemma 2.3 implies the next simple observation.

Claim 3.3. $\quad 2k \leq \pi(F) \leq (2+\varepsilon)k.$

Proof. If $\pi(F) < 2k$, then F itself does not contain H_k which has a matching of size $2k$. Hence, by coloring all the edges of F in blue, we avoid both red P_3 and blue H_k (even non-induced). This contradicts the choice of F.

If F has a matching M with $|M| > (2+\varepsilon)k$, then we color the edges of M red and all other edges blue. We do not have red P_3 at all. If we have a blue induced copy H' of H_k in F, then by Lemma 2.3, at most $|M|$ vertices incident to the edges of M can belong to $V(H')$. Hence

$$\begin{aligned}|V(H') \cup V(M)| &= |V(H')| + |V(M)| - |V(H') \cap V(M)| \\ &\geq 4k + |M| > 4k + (2+\varepsilon)k = N,\end{aligned}$$

a contradiction. □

Among the sets $W \subset V(F)$ such that $\operatorname{def}(F) = o(F-W) - |W|$ (i.e., among the sets of maximum deficiency), choose a set X of the maximum cardinality. The maximality of cardinality implies that all components of the graph $F - X$ are odd. Then by Claim 3.3,

$$(2-\varepsilon)k \leq \operatorname{def}(X) \leq (2+\varepsilon)k. \tag{5}$$

Let A_1 denote the set of components of $F - X$ that are cliques and V_1 be the set of vertices in all components in A_1. Similarly, let A_2 denote the set of components of $F - X$ that are not cliques and V_2 be the set of vertices in all components in A_2. Furthermore, let $x = |X|$ and for $i = 1, 2$, let $a_i = |A_i|$ and $v_i = |V_i|$.

Since $V(F) = X \cup V_1 \cup V_2$, we have

$$x + v_1 + v_2 = (6+\varepsilon)k. \tag{6}$$

The following is the left inequality in (5) rewritten using the names of quantities at hand:

$$(2-\varepsilon)k \leq a_1 + a_2 - x. \tag{7}$$

Claim 3.4.

$$4k \leq 2x + \frac{a_2 + v_2}{2}. \tag{8}$$

Proof. Color with red all edges in components in A_1 and a maximum matching in each component in A_2. Since X is a set of maximum deficiency, every odd component of $F-X$ (and in particular every component in A_2) has a matching saturating all but one vertex. Color all other edges of F with blue. Since every component of the obtained red graph is a clique, we have no red induced P_3. Hence, by the choice of F we have an induced blue copy H' of H_k. Note that by Lemma 2.3, H' can have at most one vertex in each component in A_1. Moreover, if a P_4 has a vertex w in a component $C \in A_1$, then all neighbors of w in this P_4 are in X. Hence $V_1 \cup X$ can contain at most $2x$ vertices of H'. Again by Lemma 2.3, each component $C \in A_2$ has at most $(1+|V(C)|)/2$ vertices of H'. Since $|V(H')| = 4k$, this proves the claim. □

If we add to Equation (6) Inequality (8) multiplied by 2 and Inequality (7) multiplied by 3 , then we get

$$8k - 4\varepsilon k + v_1 \leq 3a_1 + 4a_2. \tag{9}$$

Since $a_1 \leq v_1$, (9) yields the following:

$$4k - 2\varepsilon k \leq a_1 + 2a_2. \tag{10}$$

Claim 3.5. F has an independent set T with $|T| = 4k - 2\varepsilon k$.

Proof. Compose the independent set T' by taking a vertex from each component in A_1 and taking two non-adjacent vertices from each component in A_2. Then $|T'| = a_1 + 2a_2$ and by (10), this is at least $4k - 2\varepsilon k$. Now, let T be any subset of T' of size $4k - 2\varepsilon k$. □

From now on, we fix in F an independent set T with $|T| = 4k - 2\varepsilon k$ and let $S = V(F) - T$. Note that $|S| = 2k + 3\varepsilon k$.

Claim 3.6. The size of a maximum matching in the subgraph $F(S)$ induced by S in F is at most $3\varepsilon k$.

Proof. Suppose that there is a matching M_0 in $F(S)$ with $|M_0| > 3\varepsilon k$. Color the edges of M_0 with red and the remaining edges with blue. By the choice of F, it contains a blue induced subgraph C_0 isomorphic to H_k. Since the independence number of H_k is $2k$, at most $2k$ vertices of C_0 are in T and hence at least $2k$ vertices of C_0 should be in S. But by Lemma 2.3, S contains at most $|S| - |M_0| < (2k + 3\varepsilon k) - 3\varepsilon k = 2k$ vertices of C_0, a contradiction. □

If F does not contain induced copies of H_k, then we color all its edges blue and get a coloring contradicting the choice of F. Otherwise, choose an induced copy C_1 of H_k. Denote $B_1 = S \cap V(C_1)$ and $D_1 = T \cap V(C_1)$. Since T is independent, $|B_1| \geq 2k$. Let $|B_1| = 2k + \alpha_1 k$, where $\alpha_1 \geq 0$.

Since H_k is the union of k copies of P_4, it has the unique perfect matching, containing two edges in each copy of P_4. We will call this matching *principal*. Color the edges of the principal matching in C_1 red and all other edges of F blue. By the choice of F, we still have a blue induced in F subgraph C_2 isomorphic to H_k. Similarly to above, let $B_2 = S \cap V(C_2)$, $D_2 = T \cap V(C_2)$, and $|B_2| = 2k + \alpha_2 k$, where $\alpha_2 \geq 0$.

Observe that each vertex in $V(C_1)$ is incident to a red edge. Therefore by Lemma 2.3, each vertex in $D_1 \cap D_2$, has a neighbor (using a red edge) in $|B_1 - B_2|$. This gives us

$$|D_1 \cap D_2| \le |B_1 - B_2| \le |S - B_2| = 2k + 3\varepsilon k - (2k + \alpha_2 k) = 3\varepsilon k - \alpha_2 k \quad (11)$$

and hence

$$|D_1 \cup D_2| = |D_1| + |D_2| - |D_1 \cap D_2| \ge 2k - \alpha_1 k + 2k - \alpha_2 k - (3\varepsilon - \alpha_2)k = k(4 - 3\varepsilon - \alpha_1). \quad (12)$$

Also, we note that

$$|T - (D_1 \cup D_2)| \le 4k - 2\varepsilon k - k(4 - 3\varepsilon - \alpha_1) = \varepsilon k + \alpha_1 k \quad (13)$$

and

$$|B_1 \cap B_2| \ge |B_1| + |B_2| - |S| = 2k + \alpha_1 k + 2k + \alpha_2 k - k(2 + 3\varepsilon) = (2 + \alpha_1 + \alpha_2 - 3\varepsilon)k. \quad (14)$$

Let D' be the set of vertices in $D_1 \cup D_2$ that have two neighbors in $B_1 \cap B_2$.

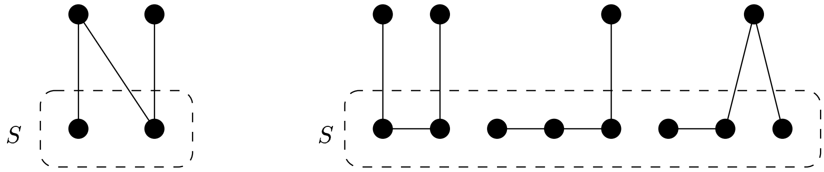
Claim 3.7. The subgraph of F induced by $D' \cup (B_1 \cap B_2)$ contains a matching M that saturates D'.

Proof. We first observe that in the graph H_k, every edge connecting $B_1 \cap B_2$ with $D_1 \cup D_2$ belongs either to $E(C_1)$ or to $E(C_2)$ and each vertex is adjacent to exactly one vertex of degree two. By the definition, each vertex in D' has two neighbors in $B_1 \cap B_2$. In particular, this means that it has two neighbors either in C_1 or in C_2. This means that each $w \in B_1 \cap B_2$ has at most two neighbors in $D_1 \cup D_2$. Hence, by Hall's Theorem, the claimed matching M exists. □

Now we color the edges of the matching M provided by Claim 3.7 with red and all other edges with blue. By the choice of F, it contains a blue induced subgraph C_3 isomorphic to H_k. We will say that a component of C_3 (which is a P_4) *is in* N*-shape*, if it has exactly two vertices in S and these vertices are not adjacent. The illustration in Fig. 4 explains why the name is used. Note that each component of C_3 that is not in N-shape has at least one edge with both ends in S. If we take an edge with both ends in S from each such component, they will form a matching of size $k - y$, where y is the number of components of C_3 in N-shape. Then Claim 3.6 yields that $y \ge k - 3\varepsilon k$.

Let $B_3 = S \cap V(C_3)$, $D_3 = T \cap V(C_3)$, and $|B_3| = 2k + \alpha_3 k$. Since $y \ge k - 3\varepsilon k$, there are at least $k - 3\varepsilon k$ vertices of D_3 that have two neighbors in B_3 (using blue edges). Let Z be the set of vertices of D_3 that have two neighbors in $B_3 \cap B_1 \cap B_2$. Since by (14),

$$|B_3 - (B_1 \cap B_2)| \le |S - (B_1 \cap B_2)| \le (2 + 3\varepsilon)k - (2 + \alpha_1 + \alpha_2 - 3\varepsilon)k = (6\varepsilon - \alpha_1 - \alpha_2)k, \quad (15)$$



P_4 in N-shape P_4's not in N-shape

Fig. 4. Possibilities for copies of P_4

we have

$$\begin{aligned}|Z| &\geq y - |B_3 - (B_1 \cap B_2)| \\ &\geq k(1-3\varepsilon) - (6\varepsilon - \alpha_1 - \alpha_2)k = k(1 - 9\varepsilon + \alpha_1 + \alpha_2). \qquad (16)\end{aligned}$$

Observe that $Z \cap (D_1 \cup D_2) = \emptyset$, because every vertex of Z has two neighbors in $B_3 \cap B_1 \cap B_2$, but every vertex of $D_1 \cup D_2$ that has two neighbors in $B_1 \cap B_2$ got one of its incident edges colored red and hence cannot be in C_3.

Thus by (13),

$$|Z| \leq |T - (D_1 \cup D_2)| \leq \varepsilon k + \alpha_1 k.$$

Comparing with (16), we get $1 - 9\varepsilon + \alpha_1 + \alpha_2 \leq \varepsilon + \alpha_1$. It follows that, $\varepsilon \geq 1/10$, a contradiction. This completes our proof of the theorem. □

Remark. Our graphs H_k are not connected. A family of connected graphs with similar properties is as follows. Let H'_k be obtained from H_k by adding a new vertex z and connecting it by an edge with an end z_i of each of the k copies of P_4 (see Fig. 5). Since H_k is an induced subgraph of H'_k, $IR(P_3, H'_k) \geq IR(P_3, H_k) \geq (6 + \frac{1}{10})k$.

For the upper bound on $IR_w(P_3, H'_k)$, consider the graph F'_k obtained from the graph F_{k+1} by adding a vertex y adjacent to a vertex y_i of degree two in every component C_i, $i = 1, \ldots, k+1$ of F_{k+1} (see Fig. 5). By construction, F'_k has $6k+7$ vertices. Let f be a red-blue coloring of the edges of F'_k. If F'_k does not have an induced red P_3, at most one edge incident with y is red. The remaining k edges yy_i are blue and by Claim 3.1 inside each of the corresponding k components C_i we have an induced P_4 starting at y_i. Thus, $IR_w(P_3, H'_k) \leq |V(F'_k)| = 6k + 7$ and

$$\lim_{k\to\infty} \sup \frac{IR(P_3, H'_k)}{IR_w(P_3, H'_k)} \geq \lim_{k\to\infty} \frac{6.1k}{6k+7} = 1 + 1/60.$$

Acknowledgement. We thank Douglas West for helpful comments.

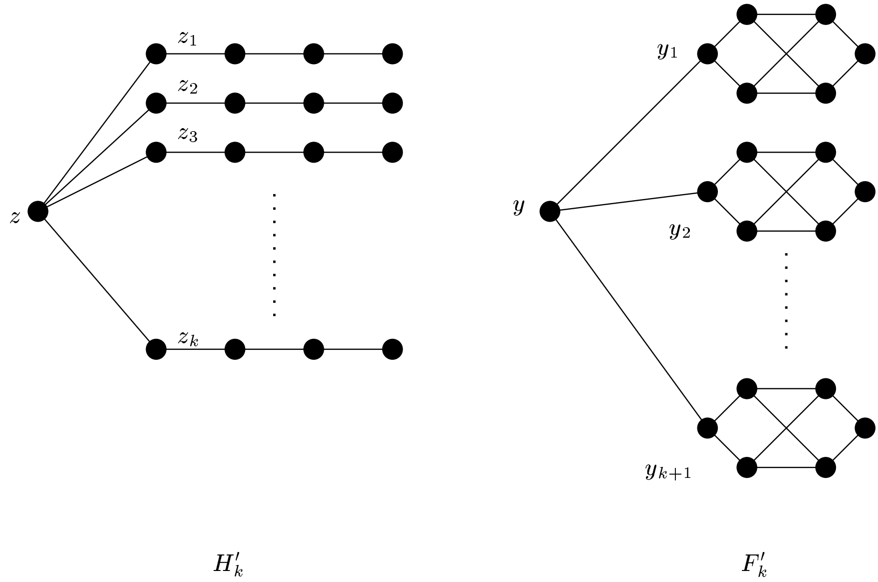


Fig. 5. The graphs H'_k and F'_k

References

[Deu75] W. Deuber, A generalization of Ramsey's theorem, in R. Rado, A. Hajnal, V. T. Sós (eds.), *Infinite and Finite Sets*, North Holland, Amsterdam, 1975.

[EHP75] P. Erdős, A. Hajnal, L. Posa, Strong embeddings of graphs into colored graphs, in R. Rado, A. Hajnal, V. T. Sós (eds.), *Infinite and Finite Sets*, North Holland, Amsterdam, 1975.

[GŁ02] I. Gorgol, T. Łuczak, On induced Ramsey numbers, *Discrete Math.*, **251** (2002), 87–96.

[HNR83] F. Harary, J. Nešetřil, V. Rödl, Generalized ramsey theory for graphs xiv: Induced ramsey numbers, in M. Fiedler (ed.), *Graphs and other combinatorical topics. Proc. of the 3rd Czechoslovak Symposium on Graph Theory*, Teubner, 1983.

[KPR98] Y. Kohayakawa, H. J. Prömel, V. Rödl, Induced Ramsey numbers, *Combinatorica* **18** (1998), 373–404.

[ŁR96] T. Łuczak, V. Rödl, On induced Ramsey numbers for graphs with bounded maximum degree, *J. of Combinatorial Theory (B)* **66** (1996), 324–333.

[HKŁ95] P. Haxell, Y. Kohayakawa, T. Łuczak, The induced size-Ramsey number of cycles, *Combinatorics, Probability and Computing* **4** (1995), 217–239.

[Neš95] J. Nešetřil, Ramsey theory, in R. Graham, M. Grötschel, L. Lovász (eds.), *Handbook of combinatorics*, Vol. 2, Elsevier, Amsterdam, 1995.

[Rödl73] V. Rödl, *A generalization of Ramsey theorem*, Ph.D. Thesis, Charles University, Prague, 1973.

[SS03] M. Schaefer, P. Shah, Induced graph Ramsey theory, *Ars Combinatoria* **66** (2003), 3–21.

On Explicit Ramsey Graphs and Estimates of the Number of Sums and Products

Pavel Pudlák*

Mathematical Insititute, Academy of Sciences of the Czech Republic, Žitná 25, 115 67 Praha 1, Czech Republic
pudlak@math.cas.cz

Summary. We give an explicit construction of a three-coloring of $K_{N,N}$ in which no $K_{r,r}$ is monochromatic for $r = N^{1/2-\varepsilon}$, where $\varepsilon > 0$ is a constant.

AMS Subject Classification. 05D10.
Keywords. Ramsey graphs, explicit constructions.

1 Introduction

In finite combinatorics there are many proofs of the existence of certain combinatorial structures which do not provide us with any explicit example of such structures. To give an explicit construction is not only a mathematical challenge, but sometimes it is the only way to determine the extremal structures for a particular question, because probabilistic existence proofs do not give us structures with matching bounds.

One of such problems is to give an explicit construction of a two-coloring of the complete bipartite graph $K_{N,N}$ such that no subgraph $K_{r,r}$ is monochromatic for some small r. (Paul Erdős asked this problem for general graphs, [CG99]; the bipartite version that we consider here is also well-known and it is believed to be harder.) It is well-known that there exist such colorings for $r = (2 + o(1)) \log_2 N$, but until recently explicit constructions were only known for $r \approx \sqrt{N}$.[1] More precisely, no proofs of such bounds were known. It has been conjectured long ago that Paley's graphs have this property for suitable finite fields, but the best bound one can prove is still only of the order of $\sqrt{N}$.

In 2003 we proposed to construct two-colorings that beat this barrier by explicitly constructing a subset of $\mathbb{F}_2^{2m}$ which has small intersections with all

* Supported by grants IAA1019401 and 1M002162080.
[1] Notice that the best lower bound is only $r = (1 - o(1)) \log_2 N$, which suggests that finding the extremal value of r may require an explicit construction.

subspaces of dimension m, see [PR04]. (We shall state this condition explicitly in the next section.) In that paper we gave a polynomial time construction of a two-coloring of $K_{N,N}$ with no monochromatic $K_{r,r}$ for $r = \sqrt{N}/2^{\sqrt{\log N}}$. Furthermore we suggested that the graphs of curves $y = x^3$ and $xy = 1$ in the field $\mathbb{F}_{2^q}$, for q prime, should give constructions beating the $\sqrt{N}$ barrier (using the natural isomorphism of the additive groups of $\mathbb{F}_{2^q}$ and $\mathbb{F}_2^q$).

Soon after that, Barak, Kindler, Shaltiel, Sudakov and Wigderson [BKSSW05] found a polynomial construction of two-colorings of $K_{N,N}$ which leave no $K_{r,r}$ monochromatic for $r = N^\varepsilon$, where ε can be chosen arbitrarily small (in fact, $\varepsilon \approx 1/\log\log N$). Their result was a breakthrough not only in the field of Ramsey graphs, but they also succeeded in constructing extractors and other gadgets needed in derandomization with much better parameters than had been known before.

The construction of Barak et al. is very complicated and uses derandomization. (Namely, one step of the construction needs structures of small size with special properties; since the size is small enough such a structure can be found by a thorough search of all structures of this size.) For applications in complexity theory this poses no problem, since from the computational point of view their construction is very effective: one can compute the color of an edge from the codes of vertices in polynomial time. Yet it seems reasonable to continue the search for more explicit constructions, even if they have worse parameters.

In this paper we give a very explicit construction of a three-coloring of $K_{N,N}$ in which no $K_{r,r}$ is monochromatic for $r = N^{1/2-\varepsilon}$, and some constant $\varepsilon > 0$. Our result is an application of the recently proved bounds on the number of sums and products in finite fields of Bourgain, Katz and Tao [BKT04]. That result is also used as the main building block of the construction of Barak et al., but in a different way.

Our aim was also to prove the conjectures about curves $y = x^3$ and $xy = 1$ mentioned above. We have succeeded only partially, namely we can prove the corresponding statement for $y = x^2$ and for fields of characteristics different from 2. For fields of the characteristic 2 this curve is not good for our purpose. Since the number of colors is the size of the prime subfield, the smallest number of colors that we can get is 3. The most recent results of Bourgain [Bou06$^+$a, Bou06$^+$b] in this area seem also to confirm our conjecture about curves $y = x^3$ and $xy = 1$. In [Bou06$^+$b] Bourgain defined three explicit two-colorings of $K_{N,N}$ in which no $K_{r,r}$ is monochromatic for $r = N^{1/2-\varepsilon}$, for some $\varepsilon > 0$.

2 The Three-coloring of $K_{N,N}$

Let F be a field. Let $S \subseteq F^n$. We define a coloring γ of the complete bipartite graph $S' \times S''$, where $S' = \{1\} \times S$ and $S'' = \{2\} \times S$, by the formula

$$\gamma((1,u),(2,v)) = \langle u,v \rangle,$$

where $\langle u,v \rangle = \sum_{i=1}^{n} u_i v_i$ is the inner product in F.

In plain words, we take two copies of the subset S of the vector space and color every pair of vertices, with u in one copy and v in the other copy, by the element of the field F equal to the inner product of the two vertices. Thus if $N = |S|$ and $c = |F|$, γ is a coloring of $K_{N,N}$ by c colors.

In [PR04] we proved the following simple proposition only for the two-element field, but the proof is completely general. Hint: think of A as a set of equations and B as a set of solutions.

Proposition 2.1. *Suppose every vector space $V \subseteq F^n$ of dimension $\lfloor (n+1)/2 \rfloor$ intersects S in less than r elements. Then no complete bipartite subgraph $K_{r,r}$ is monochromatic with respect to γ, ie., for no two subsets $A \subseteq S_1$, $B \subseteq S_2$, $|A| = |B| = r$ the value of $\gamma(a,b)$ is the same for all $a \in A$ and $b \in B$.*

We shall consider the following construction of S. Let $p > 2$ be a prime and $n = 2q$. Put

$$S_{p,q} = \{(x,x^2);\ x \in \mathbb{F}_{p^q}\}.$$

In order to define a coloring on the product of two copies of $S_{p,q}$, think of $\mathbb{F}_{p^q}$ as a q-dimensional vector space over $\mathbb{F}_p$. Thus $S_{p,q} \subseteq \mathbb{F}_p^n$ and we can define $\gamma_{p,q}$ using the scalar product in F_p. Hence $\gamma_{p,q}$ is a coloring of $K_{N,N}$, $N = p^q$, by p colors.

Our main result is the following theorem.

Theorem 2.2. *For every prime $p > 2$ there exists $\varepsilon > 0$ such that for every sufficiently large prime q, the coloring $\gamma_{p,q}$ of $K_{N,N}$ has no monochromatic subgraph $K_{r,r}$ for $r > N^{1/2-\varepsilon}$.*

We shall first explain the main ideas of the proof. By Proposition 2.1 it suffices to estimate from above the size of the intersections $S_{p,q} \cap V$ for subspaces of $\mathbb{F}_p^n$ of dimension $q = n/2$. Notice that if we view $\mathbb{F}_p^n$ as $\mathbb{F}_{p^q}^2$ then it is an affine plain and $S_{p,q}$ is a parabola in it. A line intersects a parabola in at most two points. A line in $\mathbb{F}_{p^q}^2$ is a q-dimensional subspace if we view it in $\mathbb{F}_p^n$. There are many more subspaces of this dimension, but our hope is the their intersections with $S_{p,q}$ are also small.

Now instead of estimating the intersections, we shall consider subsets S of $S_{p,q}$ of certain size and show that they span dimension bigger than q. This we also do not do directly. We first estimate the size of $S+S+S$ and take the logarithm. (We could take more than three terms, but the gain would not be significant.) For estimating the size of $S+S+S$ we use some standard techniques and their recent extensions to finite fields.

Now we proceed with a formal proof. Let the field $\mathbb{F}_p$ be fixed for the proof of this theorem. The following is a finite field version of Theorem 1 of Elekes, Nathanson and Ruzsa [ENR99] originally proved for the real numbers and $S \subseteq \{(x,f(x));\ x \in \mathbb{R}\}$ for every strictly convex function f in place of x^2.

Lemma 2.3. *For every $\alpha > 0$ there exist $\varepsilon_0, \varepsilon_1 > 0$ such that for every sufficiently large prime q, every subset $S \subseteq S_{p,q}$ and every set $T \subseteq \mathbb{F}_p^{2q}$, if $p^{\alpha q} \leq |T| \leq p^{(2-\alpha)q}$ then*

$$|S + T| \geq \varepsilon_0 |S| \cdot |T|^{1/2+\varepsilon_1}.$$

We shall first prove the theorem using this lemma. Let V be a vector subspace of $\mathbb{F}_p^{2q}$ of dimension $q + 1$. Put $S = S_{p,q} \cap V$ and $T = S + S$. Then $|T| \geq \binom{|S|+1}{2}$, since the pair $(x + y, x^2 + y^2)$ uniquely determines the set $\{x, y\}$. We can apply the previous lemma to T, since $T \subseteq V$, hence $|T| \leq p^{q+1}$. According to the lemma we thus have

$$|S + S + S| \geq \varepsilon_0 |S| \cdot \binom{|S|+1}{2}^{1/2+\varepsilon_1} \geq \varepsilon_0 |S|^{2+\varepsilon_1}/2.$$

Hence the dimension of the vector space spanned by S is at least $\log_p(\varepsilon_0 |S|^{2+\varepsilon_1}/2)$. This must be at most the dimension of V, hence

$$\log_p(\varepsilon_0 |S|^{2+\varepsilon_1}/2) \leq q + 1,$$

from which we get

$$|S| \leq \left(2\varepsilon_0^{-1} p^{q+1}\right)^{\frac{1}{2+\varepsilon_1}} \leq p^{(\frac{1}{2}-\varepsilon)q}$$

for some $\varepsilon > 0$. □

To prove Lemma 2.3 we shall use the following an estimate on the number of incidences of points and lines in a finite plane proved by Bourgain, Katz and Tao in [BKT04] as Theorem 6.2.

Theorem 2.4. *Let $0 < \alpha < 2$. Then there exist constants $0 < \beta < 1$, $\varepsilon_2 > 0$ and C such that for every finite field F, set of points P and set of lines L in the projective plane over F, if $|P|, |L| \leq N = |F|^\alpha$ and F does not contain a subfield of size bigger than $|F|^\beta$, then*

$$I_{P,L} \leq CN^{3/2-\varepsilon_2},$$

where $I_{P,L} = |\{(p, l) \in P \times L;\ p \in l\}|$ denotes the number of incidences.

In [BKT04] the theorem is proven only for prime fields and a stronger statement which implies the theorem above is stated without a proof. However it is easy to verify the stronger statement by inspecting the proof in [BKT04]. In fact, there is only one step in their proof that needs the assumption that the field is prime, which is Lemma 4.1. One can immediately see that the proof of this lemma works perfectly if we only assume that the field does not contain a large subfield.

We shall need an estimate for the case when the number of lines and the number of points is different.

Corollary 2.5. *For every $0 < \alpha' < \alpha < 2$, there exist constants $0 < \beta < 1$, $\varepsilon_3 > 0$ and C' such that for every F, P, L, if $|F|^{\alpha'} \leq |L| \leq |P| \leq |F|^{\alpha}$ and F does not contain a subfield of size bigger than $|F|^{\beta}$, then*

$$I_{P,L} \leq C'|P| \cdot |L|^{\frac{1}{2}-\varepsilon_3}.$$

Proof. Let $P' \subseteq P$ be a random subset of P of size $|L|$. Then the expected value of the number of incidences $I_{P',L}$ is $I_{P,L}|L|/|P|$. Thus there exists P' such that $I_{P',L} \geq I_{P,L}|P'|/|P| = I_{P,L}|L|/|P|$. Applying the theorem to P' and L, we get

$$I_{P,L}|L|/|P| \leq I_{P',L} \leq C'|L|^{3/2-\varepsilon_3},$$

for some $\varepsilon_3 > 0$ and C', whence we get the statement of the corollary. ☐

Now we shall prove Lemma 2.3. Let $S \subseteq S_{p,q}$ and $T \subseteq \mathbb{F}_p^{2q}$ be given. Put $Q = \{S_{p,q} + t;\ t \in T\}$. We think of $S_{p,q}$ as a parabola in the affine plane and Q as the set of all shifts of this parabola by vectors $t \in T$. Put $P = S + T$. So P is a set of points on parabolas Q. We want to use the estimate on the number of incidences in Corollary 2.5. The corollary speaks only about sets of lines, but we can show that a suitable one-to-one transformation maps our parabolas on lines. This mapping is defined by $(u, v) \mapsto (u, v - u^2)$, and it maps the parabola $S_{p,q} + (a, b)$ onto the line

$$\{(x + a, 2ax - a^2 + b);\ x \in \mathbb{F}_{p^q}\}.$$

The number of incidences is $|S| \cdot |T|$, since we have $|T|$ parabolas in Q, and on each parabola $Q + t$ we have $|S|$ points, namely the points $S + t$. Thus by Corollary 2.5, we have

$$|S| \cdot |T| = I_{P,Q} \leq C'|P| \cdot |Q|^{\frac{1}{2}-\varepsilon_3} = C'|S + T| \cdot |T|^{\frac{1}{2}-\varepsilon_3},$$

whence Lemma 2.3 follows. ☐

Proposition 2.6. *For $p > 2$ prime and q arbitrary positive integer, $K_{N,N}$ colored by $\gamma_{p,q}$ contains a monochromatic subgraph $K_{r,r}$ for $r = \varepsilon_4 N^{1/4}$, for some $\varepsilon_4 > 0$.*

Proof. Represent the elements of $\mathbb{F}_{p^q}$ as polynomials modulo an irreducible polynomial of degree q over $\mathbb{F}_p$. Let A be the set of all polynomials of degree less than $q/4$ and let B be the set of all polynomials that have nonzero coefficients only at terms of degree n for $q/4 \leq n < q/2$. Then the polynomials that represent the squares of elements of A are polynomials of degree less than $q/2$ and the polynomials that represent the squares of elements of B are polynomials that have nonzero coefficients at terms of degree n for $q/2 \leq n < q$. Hence the scalar product of every pair $a \in A$ and $b \in B$ is zero. ☐

We do not know other monochromatic subgraphs $K_{r,r}$.

3 Concluding Remarks

We observe that our construction possess a symmetry property which implies a slightly stronger result than stated above. We construct a three-coloring of K_N such that for some $\varepsilon > 0$ independent of N the coloring has the following property. There are no two subsets of vertices X and Y of size at least $N^{1/2-\varepsilon}$ (disjoint or not disjoint) such that all edges between X and Y have the same color.

The most interesting open problem related to our result is whether we can get a two-coloring in such a way. If $p = 2$, then we cannot use $S_{p,q}$, because x^2 is an additive function in fields of characteristic 2, thus $S_{p,q}$ is a linear subspace of $\mathbb{F}_2^n$ and $\gamma_{2,q}$ is 0 for all edges. In [PR04] we proposed to use

$$\{(x, x^{-1});\ x \in \mathbb{F}_{2^q}\}, \quad \text{and} \quad \{(x, x^3);\ x \in \mathbb{F}_{2^q}\}.$$

We conjecture that the same statement as our Theorem 2.2 holds for $p = 2$ and the sets above. One could prove it in the same way if we had a generalization of the bound on the number of incidences of points and lines (Theorem 2.4) to hyperbolas and cubics. The corresponding result has been proven in the Euclidean plane for a much broader class of curves. Let us note that the graphs defined using the curve $y = x^3$ contain a monochromatic $K_{r,r}$ for $r = \varepsilon_5 N^{1/6}$, for some $\varepsilon_5 > 0$ (the proof is the same as in Proposition 2.6). For $y = x^{-1}$ we do not have any such result and we conjecture that they do not contain $K_{N^\varepsilon, N^\varepsilon}$ for any $\varepsilon > 0$.

The bound on the number of incidences in a finite plane is an application of the lower bound on the number of sums and products

$$|A + A| \cdot |A \cdot A| \geq \delta |A|^{2+\varepsilon}$$

for some constants $\delta, \varepsilon > 0$, provided that A is not too small or too big in the finite field. (The first restriction has been removed in a paper of Konyagin [Kon03] at least for prime fields.) The transformation of the parametrized set of parabolas to lines used above can also be applied to prove a similar estimate

$$|A + A| \cdot |A^2 + A^2| \geq \delta |A|^{2+\varepsilon}.$$

in finite fields. For hyperbolas, ie.,

$$|A + A| \cdot |A^{-1} + A^{-1}| \geq \delta |A|^{2+\varepsilon},$$

this was recently proved by Bourgain [Bou06⁺a]. For cubics such a bound is not known, but it is very likely to be true. In [Bou06⁺b] Bourgain proved another bound related to these problems

$$\{xy(x + y);\ x, y \in A\}| \geq \delta |A|^{1+\varepsilon}.$$

For comparison with our construction we state the definition of one of the two-colorings of Bourgain [Bou06⁺b] mentioned in the introduction. For a prime p, $f : \mathbb{F}_p \times \mathbb{F}_p \to \{\pm 1\}$ is defined by

$$f(x,y) = \operatorname{sgn}\sin\frac{2\pi}{p}(xy + x^2y^2),$$

with the convention that $\operatorname{sgn} 0 = 1$. Bourgain proved a stronger property of f, namely that f defines a *two source randomness extractor*, which means that for every subgraph $K_{r,r}$, $r = N^{1/2-\varepsilon}$ the discrepancy between the number of edges colored 1 and edges colored -1 is $p^{-\gamma}r^2$ for some $\gamma > 0$.

Acknowledgement. I would like to thank to Jiří Matoušek for explaining me a proof of Elekes and to Jiří Sgall for suggesting an idea for Proposition 2.6. I also appreciate helpful remarks of the anonymous referees.

References

[BKSSW05] B. Barak, G. Kindler, R. Shaltiel, B. Sudakov and A. Wigderson, Simulating Independence: New Constructions of Condensers, Ramsey Graphs, Dispersers, and Extractors, *ACM Symp. on Theory of Computing* 2005, 1–10.

[Bou06+a] J. Bourgain, More on the sum-product phenomenon in prime fields and its applications, preprint 2005, to appear in *IJNT.*

[Bou06+b] J. Bourgain, On the construction of affine extractors, preprint 2005.

[BKT04] J. Bourgain, N. Katz and T. Tao, A sum-product estimate in finite fields, and applications, *GAFA*, 14:1 (2004), 27–57.

[CG99] F. Chung and R. Graham, *Erdős on Graphs, His Legacy of Unsolved Problems.* A K Peters, 1999.

[ENR99] G. Elekes, M. B. Nathanson and I. Z. Ruzsa, Convexity and Sumsets, *J. of Number Theory* **83** (1999), 194–201.

[Kon03] S. V. Konyagin, A sum-product estimate in fields of prime order, arXiv:math.CO/0301343 v2, 2003.

[PR04] P. Pudlák and V. Rödl, Pseudorandom sets and explicit constructions of Ramsey graphs, *Quaderni di Matematica*, vol. 13, Seconda Università di Napoli, 2004, 327–46.

Part III

Graphs and Hypergraphs

Hereditary Properties of Ordered Graphs

József Balogh[1], Béla Bollobás[2], and Robert Morris[3]*

[1] Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL 61801, USA
jobal@math.uiuc.edu

[2] Department of Mathematical Sciences, The University of Memphis, Memphis, TN 38152, USA, and Trinity College, Cambridge CB2 1TQ, England
bollobas@msci.memphis.edu

[3] Department of Mathematical Sciences, The University of Memphis, Memphis, TN 38152, USA
rdmorrs1@memphis.edu

Summary. An ordered graph is a graph together with a linear order on its vertices. A hereditary property of ordered graphs is a collection of ordered graphs closed under taking order-preserving isomorphisms of the vertex set, and order-preserving induced subgraphs. If $\mathcal{P}$ is a hereditary property of ordered graphs, then $\mathcal{P}_n$ denotes the collection $\{G \in \mathcal{P} : V(G) = [n]\}$, and the function $n \mapsto |\mathcal{P}_n|$ is called the speed of $\mathcal{P}$.

The possible speeds of a hereditary property of labelled graphs have been extensively studied (see [BBW00] and [Bol98] for example), and more recently hereditary properties of other combinatorial structures, such as oriented graphs ([AS00], [BBM06$^+$c]), posets ([BBM06$^+$a], [BGP99]), words ([BB05], [QZ04]) and permutations ([KK03], [MT04]), have also attracted attention. Properties of ordered graphs generalize properties of both labelled graphs and permutations.

In this paper we determine the possible speeds of a hereditary property of ordered graphs, up to the speed 2^{n-1}. In particular, we prove that there exists a jump from polynomial speed to speed F_n, the Fibonacci numbers, and that there exists an infinite sequence of subsequent jumps, from $p(n)F_{n,k}$ to $F_{n,k+1}$ (where $p(n)$ is a polynomial and $F_{n,k}$ are the generalized Fibonacci numbers) converging to 2^{n-1}. Our results generalize a theorem of Kaiser and Klazar [KK03], who proved that the same jumps occur for hereditary properties of permutations.

AMS Subject Classification. 05C99.
Keywords. Ordered graphs, hereditary property, speed.

* The first author was supported during this research by OTKA grant T049398 and NSF grant DMS-0302804, the second by NSF grant ITR 0225610, and the third by a Van Vleet Memorial Doctoral Fellowship.

1 Introduction

In this paper we shall determine the possible speeds of a hereditary property of ordered graphs, up to the speed 2^{n-1}. In particular, we shall prove that there is a jump from polynomial speed to speed F_n, the Fibonacci numbers, and that there exists an infinite sequence of subsequent jumps converging to 2^{n-1}. Our results generalize a theorem of Kaiser and Klazar [KK03], who proved that the same jumps occur for hereditary properties of permutations. We begin by making the definitions necessary in order to state our main result.

An *ordered graph* is a graph together with a linear order on its vertices. As a convention, we shall assume that if G is an ordered graph of order n, then $V(G) = [n]$, where $i < j$ as vertices of G if $i < j$ in $\mathbb{N}$. A collection of ordered graphs is called a *property* if it is closed under order-preserving isomorphisms of the vertex set. Given ordered graphs G and H, we say that G is an *induced ordered subgraph* of H (and write $G \leqslant H$) if there exists an injective, order-preserving map $\phi : V(G) \to V(H)$ such that $ij \in E(G)$ if and only if $\phi(i)\phi(j) \in E(H)$. A property of ordered graphs $\mathcal{P}$ is called *hereditary* if it is closed under taking induced ordered subgraphs. In this paper 'subgraph' will always mean 'induced ordered subgraph', unless otherwise stated. Given a property of ordered graphs, $\mathcal{P}$, we write $\mathcal{P}_n$ for the collection of ordered graphs in $\mathcal{P}$ with vertex set $[n]$. The *speed* of $\mathcal{P}$ is simply the function $n \mapsto |\mathcal{P}_n|$. Analogous definitions can be made for other combinatorial structures (e.g., graphs, posets, permutations).

We are interested in the (surprising) phenomenon, observed for hereditary properties of various types of structure (see for example [BBW00], [BGP99], [MT04]) that the speeds of such a property are far from arbitrary. More precisely, there often exists a family $\mathcal{F}$ of functions $f : \mathbb{N} \to \mathbb{N}$ and another function $F : \mathbb{N} \to \mathbb{N}$ with $F(n)$ *much* larger than $f(n)$ for every $f \in \mathcal{F}$, such that if for each $f \in \mathcal{F}$ the speed is infinitely often larger than $f(n)$, then it is also larger than $F(n)$ for every $n \in \mathbb{N}$. Putting it concisely: the speed *jumps* from $\mathcal{F}$ to F.

We can now state our main result. Let $F_{n,k}$ denote the n^{th} generalized Fibonacci number of order k, defined by $F_{n,k} = 0$ if $n < 0$, $F_{0,k} = 1$ and $F_{n,k} = F_{n-1,k} + F_{n-2,k} + \ldots + F_{n-k,k}$ for every $n \geqslant 1$. We shall sometimes write F_n for $F_{n,2}$.

Theorem 1.1. *If $\mathcal{P}$ is a hereditary property of ordered graphs, then one of the following assertions holds.*

(a) *$|\mathcal{P}_n|$ is bounded, and there exist $M, N \in \mathbb{N}$ such that $|\mathcal{P}_n| = M$ for every $n \geqslant N$.*
(b) *$|\mathcal{P}_n|$ is a polynomial. There exist $k \in \mathbb{N}$ and integers $a_0, \ldots, a_k$ such that, $|\mathcal{P}_n| = \sum_{i=0}^{k} a_i \binom{n}{i}$ for all sufficiently large n, and $|\mathcal{P}_n| \geqslant n$ for every $n \in \mathbb{N}$.*
(c) *$F_{n,k} \leqslant |\mathcal{P}_n| \leqslant p(n)F_{n,k}$ for every $n \in \mathbb{N}$, for some $2 \leqslant k \in \mathbb{N}$ and some polynomial p, so in particular $|\mathcal{P}_n|$ is exponential.*
(d) *$|\mathcal{P}_n| \geqslant 2^{n-1}$ for every $n \in \mathbb{N}$.*

We remark that our theorem is inspired in part by the work of Kaiser and Klazar [KK03], who proved an identical theorem for hereditary properties of permutations. In fact Theorem 1.1 is a generalization of their result, since every hereditary property of permutations $\mathcal{Q}$ may be thought of as a hereditary property of ordered graphs $\mathcal{P}(\mathcal{Q})$ in the following way. For each $n \in \mathbb{N}$ and $\pi \in \mathcal{Q}_n$, let $G(\pi)$ be the ordered graph with vertex set $[n]$, and with edge set $\{ij : \text{the order of the vertices } i \text{ and } j \text{ is reversed by } \pi\}$. Let $\mathcal{P}(\mathcal{Q}) = \{G(\pi) : \pi \in \mathcal{Q}\}$. It is easy to see that $\mathcal{P}(\mathcal{Q})$ is hereditary, and that $|\mathcal{P}(\mathcal{Q})_n| = |\mathcal{Q}_n|$. In fact, writing Π for the collection of all permutations, one can give an even simpler description of the property $\mathcal{P}(\Pi)$. Let H_1 denote the ordered graph on vertex set [3] and with edge set $\{12, 23\}$, and H_2 that with edge set $\{13\}$. Then $\mathcal{P}(\Pi) = \{G : H_1 \not\leqslant G \text{ and } H_2 \not\leqslant G\}$, where $H_i \not\leqslant G$ means that H_i is not an induced ordered subgraph of G. Hence the theorem of Kaiser and Klazar is Theorem 1.1 in the case that $H_1, H_2 \notin \mathcal{P}$.

There are other interesting special cases of Theorem 1.1. For example, let $\mathcal{G}$ be a hereditary property of oriented graphs, and let $\mathcal{G}^n_{mon}$ denote the collection of pairs (G, ϕ), where $G \in \mathcal{G}$, $|G| = n$ and $\phi : [n] \leftrightarrow V(G)$ is a monotone labelling of the vertices of G, i.e., if $x \to y$ in G then $\phi(x) < \phi(y)$. Since each monotone labelling may be thought of as an ordering, there is a hereditary property of ordered graphs $\mathcal{P}$ such that $|\mathcal{G}^n_{mon}| = |\mathcal{P}_n|$ for every $n \in \mathbb{N}$. Hence the speed $n \mapsto |\mathcal{G}^n_{mon}|$ satisfies the conclusion of Theorem 1.1.

Similarly, let $\mathcal{R}$ be a hereditary property of posets, and let $\mathcal{P}^n_{lin}$ denote the collection of pairs (P, ψ), where $P \in \mathcal{R}$, $|P| = n$, and ψ is a linear extension of P, i.e., a monotone labelling of the elements of P. Each pair (P, ψ) may be thought of as a transitive monotone-labelled oriented graph, so the speed $n \mapsto |\mathcal{R}^n_{lin}|$ also satisfies the conclusion of Theorem 1.1. We suspect that our list of interesting special cases is not exhaustive.

A property of *graphs* is a collection of (unlabelled) graphs closed under isomorphism, and a property of graphs is hereditary if it is closed under taking (non-ordered) induced subgraphs. Given a property of graphs, $\mathcal{P}$, we write $\mathcal{P}^n$ for the collection of labelled graphs of order n in $\mathcal{P}$ (i.e., the collection of non-isomorphic pairs (G, ϕ), where $G \in \mathcal{P}$, $|G| = n$ and $\phi : [n] \leftrightarrow V(G)$ is a labelling of the vertices of G), and call the function $n \mapsto |\mathcal{P}^n|$ the labelled speed of $\mathcal{P}$. The labelled speed of a property of graphs was introduced by Erdős [Erd64] in 1964, and subsequently studied by Erdős, Kleitman and Rothschild [EKR76], Erdős, Frankl and Rödl [EFR86], Kolaitis, Prömel and Rothschild [KPR87], and Prömel and Steger [PS92], [PS96], [PS96], amongst others, though always in the special case where only a single graph is forbidden. The study of the possible speeds of a *general* hereditary property of labelled graphs was initiated by Scheinerman and Zito [SZ94] in 1994. They were the first to study speeds below n^n, and proved that for such properties the speeds all lie in a few fairly narrow ranges. A little later, considerably stronger results were proved by Balogh, Bollobás and Weinreich [BBW00], [BBW06+]. In the range 2^{cn^2}, the main results were proved by Alekseev [Ale93], by Bollobás and Thomason [BT95], [BT97], and by Prömel and Steger [PS96].

Putting all these results together, one obtains the following powerful theorem. Here B_n denotes the n^{th} Bell number, the number of partitions of $[n]$.

Theorem A. *Let $\mathcal{P}$ be a hereditary property of graphs. Then one of the following holds.*

(a) $|\mathcal{P}^n| = \sum_{i=1}^{k} p_i(n) i^n$ *for every $n \geqslant N$, for some $N, k \in \mathbb{N}$, and some collection $p_1(n), \ldots, p_k(n)$ of polynomials.*

(b) $|\mathcal{P}^n| = n^{(1-1/k+o(1))n}$ *as $n \to \infty$, for some $2 \leqslant k \in \mathbb{N}$.*

(c) $n^{(1+o(1))n} = B_n \leqslant |\mathcal{P}^n| \leqslant 2^{o(n^2)}$ *as $n \to \infty$.*

(d) $|\mathcal{P}^n| = 2^{(1-1/k+o(1))\binom{n}{2}}$*, as $n \to \infty$, for some $2 \leqslant k \in \mathbb{N}$.*

(e) $|\mathcal{P}^n| = 2^{\binom{n}{2}}$ *for every $n \in \mathbb{N}$.*

Given a property of graphs $\mathcal{G}$, one can define a property of ordered graphs $\mathcal{P}(\mathcal{G})$ by taking every possible ordering on the vertex set of each graph in $\mathcal{G}$. Note that now $|\mathcal{P}(\mathcal{G})_n| = |\mathcal{G}^n|$, so the speed of a property of ordered graphs is also a generalization of the labelled speed of a property of graphs. Using this idea, we can easily deduce the following theorem from Theorem A.

Theorem 1.2. *Let $\mathcal{P}$ be a hereditary property of ordered graphs. Either $|\mathcal{P}_n| = 2^{o(n^2)}$ as $n \to \infty$, or $|\mathcal{P}_n| = 2^{(1-1/k+o(1))\binom{n}{2}}$ for some $2 \leqslant k \in \mathbb{N}$.*

Proof. Given a hereditary property of ordered graphs, $\mathcal{P}$, we can naturally associate a property of graphs $\mathcal{G}$ with $\mathcal{P}$, by identifying isomorphic graphs with different linear orders. Since $\mathcal{P}$ is hereditary, so is $\mathcal{G}$. Also, since each linear order may be thought of as a labelling, and there are at most $n!$ different labellings of a graph in $\mathcal{G}_n$, we have

$$|\mathcal{G}_n| \leqslant |\mathcal{P}_n| \leqslant |\mathcal{G}^n| \leqslant n! \cdot |\mathcal{G}_n| \tag{1}$$

for every $n \in \mathbb{N}$. So if $|\mathcal{P}_n| \geqslant 2^{cn^2}$ for some $c > 0$ and for infinitely many n, then also $|\mathcal{G}^n| \geqslant 2^{cn^2}$ for these n. Hence, by Theorem A, $|\mathcal{G}^n| = 2^{(1-1/k+o(1))\binom{n}{2}}$ for some $2 \leqslant k \in \mathbb{N}$.

Now, by (1), we also have $|\mathcal{G}_n| = 2^{(1-1/k+o(1))\binom{n}{2}}$, since $n! = 2^{o(n^2)}$, and so $|\mathcal{P}_n| = 2^{(1-1/k+o(1))\binom{n}{2}}$, as claimed. □

Theorems 1.1 and 1.2 determine the possible speeds of a hereditary property of ordered graphs below 2^{n-1} and above 2^{cn^2}, but in the large range in between many questions remain. In [BBM06⁺b] the current authors conjectured that for such a property $\mathcal{P}$ either $|\mathcal{P}_n| < c^n$ for some constant c, or $|\mathcal{P}_n| \geqslant \sum_{k=0}^{\lfloor n/2 \rfloor} \binom{n}{2k} k!$ for every $n \in \mathbb{N}$. They also proved several special cases of the conjecture (some of which were proved independently by Klazar and

Marcus [KM06+]), each of which generalizes the well-known Stanley-Wilf conjecture, which was recently proved by the combined results of Klazar [Kla00] and Marcus and Tardos [MT04].

We shall discuss these and other open questions in greater length in Section 8, but let us now return to the proof of Theorem 1.1. The proof will require some notation, and for convenience we shall give a portion of it here, before we begin.

Let G be an ordered graph with $V(G) = [n]$. The *length* of the edge $ij \in E(G)$ is $|i - j|$, and G is *ℓ-complete* if it has all edges of length at least ℓ, and *ℓ-empty* if it has none. If $x \in [n]$, and $\ell \in \mathbb{N}$, then let $N_\ell(x) = [x - \ell + 1, x + \ell - 1]$ (where $[a, b] = \{k \in \mathbb{N} : a \leqslant k \leqslant b\}$). Let $\Gamma(x)$ denote the set of neighbours of x in G. We say x and y are *ℓ-homogeneous* (and write $x \sim_\ell y$) if $\Gamma(x) \setminus (N_\ell(x) \cup N_\ell(y)) = \Gamma(y) \setminus (N_\ell(x) \cup N_\ell(y))$, and say that $B \subset V(G)$ is an *ℓ-homogeneous block* if it is a set of consecutive vertices such that $x \sim_\ell y$ for every $x, y \in B$. Note that if B is a ℓ-homogeneous block then $G[B]$ is ℓ-complete or ℓ-empty. If B is a maximal 1-homogeneous block, we say simply that it is a *homogeneous block*. Note that $\sim_1$ is an equivalence relation, so the homogeneous blocks of G are determined uniquely.

Let $A, B \subset [n]$. We shall write $A < B$ if $a < b$ for every $a \in A$ and $b \in B$, and $A \sim_\ell B$ if $a \sim_\ell b$ for every $a \in A$ and $b \in B$. If G is an ordered graph, and $A, B \subset V(G)$ with $A < B$, let $G[A]$ denote the ordered graph induced by the set A, and let $G[A, B]$ denote the bipartite ordered graph induced by the edges between A and B. We shall write $G - A$ for $G[[n] \setminus A]$, and $G[a, b]$ for $G[[a, b]]$ if $a, b \in \mathbb{N}$. Finally, $G \leqslant H$ will mean (as above) that G is an induced ordered subgraph of H.

The rest of the paper is organised as follows. In Section 2 we shall prove the key lemma in the proof of Theorem 1.1; in Section 3 we shall prove that the existence of certain structures in $\mathcal{P}$ implies that the speed is at least 2^{n-1}; in Section 4 we shall prove the jump from polynomial speed to F_n; in Section 5 we shall prove various lemmas about ℓ-empty ordered graphs; in Section 6 we shall deduce the structure of a property with speed $p(n)F_{n,k}$; in Section 7 we shall prove Theorem 1.1; and in Section 8 we shall discuss some further problems, including the possible exponential speeds above 2^{n-1}.

2 The Key Lemma

We start by defining, for each pair $k, \ell \in \mathbb{N}$, ten basic structures. The structures come in four flavours.

Type 1: there are vertices y and $x_1 < \ldots < x_{2k}$ in G such that $y < x_1$ or $y > x_{2k}$, and for $1 \leqslant i < 2k$, $yx_i \in E(G)$ iff $yx_{i+1} \notin E(G)$.

Type $2(a)$: there are vertices $x_1 < \ldots < x_{2k} < y_1 < \ldots < y_{2k}$ in G such that $x_i y_i \in E(G)$ iff $x_{i+1} y_{i+1} \notin E(G)$.

Type $2(b)$: there are vertices $x_1 < \ldots < x_{2k} < y_{2k} < \ldots < y_1$ in G such that

$x_i y_i \in E(G)$ iff $x_{i+1} y_{i+1} \notin E(G)$.

Type 3: there are vertices $x_1 < z_{1,1} < \ldots < z_{1,\ell-1} < y_1 < x_2 < z_{2,1} < \ldots < z_{2,\ell-1} < y_2 < x_3 < \ldots < y_{2k-1} < x_{2k} < z_{2k,1} < \ldots < z_{2k,\ell-1} < y_{2k}$ in G such that $x_i y_i \in E(G)$ iff $x_{i+1} y_{i+1} \notin E(G)$.

Note that there are four different structures of Type 1, and two each of Types 2(a), 2(b) and 3. We refer to these as k-structures of Type 1 and 2, and (k,ℓ)-structures of Type 3 (throughout we shall say "Type 2" when we mean "Type 2(a) or Type 2(b)"). They are not graphs, but sub-structures contained in graphs: instead of saying that "a structure of Type i occurs in $\mathcal{P}$" it would be more precise to say that "there is a graph $G \in \mathcal{P}$ admitting a structure of Type i". However, for smoothness of presentation, we sometimes handle them as graphs.

The key lemma in the proof of Theorem 1.1 will be the following.

Lemma 2.1. *Let $k, \ell \in \mathbb{N}$, and G be any ordered graph. If G contains no k-structure of Type 1 or 2, and no (k,ℓ)-structure of Type 3, then the vertices of G may be partitioned into blocks $B_1 < \ldots < B_m$, with $m \leqslant 256k^4$, and each block ℓ-homogeneous.*

Proof. Let $k, \ell \in \mathbb{N}$, and G be an ordered graph with vertex set $[n]$. Suppose that no k-structure of Type 1 or 2, and no (k,ℓ)-structure of Type 3 occurs in G, and suppose without loss of generality that $(1, \ell+1) \in E(G)$. Let $i_0 = 0$, $i_1 = \ell + 1$, and let i_2 be minimal under the condition that it is an endpoint of a non-edge ji_2, with $i_1 = \ell + 1 < j \leqslant i_2 - \ell$. Now fixing i_2, let i_3 be the minimal number under the condition that it is an endpoint of an edge ji_3, with $i_2 < j \leqslant i_3 - \ell$. If no such edge / non-edge exists at stage t, then set $i_t = n + 1$ and stop. Continuing in this way, the sequence $\{i_0, \ldots, i_t\}$ may be defined, but it may not continue further than $t = 2k$, otherwise a (k,ℓ)-structure of Type 3 would appear as a sub-structure. For each $j \in [t]$, the graph spanned by $[i_{j-1} + 1, i_j - 1]$ is either ℓ-complete or ℓ-empty, depending on the parity of j. This means that the vertex set of G can be partitioned into at most $2k - 1$ vertices and at most $2k$ blocks $A_1 < \ldots < A_t$ of consecutive vertices, where the blocks span ℓ-complete or ℓ-empty graphs.

Let $j \in [t]$ and consider the block A_j. Let $A_j = [u_j, v_j]$, and let $S_j = \{s \in [u_j, v_j - 1] : \Gamma(s) \setminus A_j \neq \Gamma(s+1) \setminus A_j\}$ be the set of vertices 'separated' from the next vertex to the right by a vertex outside A_j. We shall show that $|S_j| < 64k^3$.

For each $s \in S_j$, choose a vertex $w = w(s) \in [n] \setminus A_j$ such that $sw \in E(G)$ but $(s+1)w \notin E(G)$, or vice versa. Let $T_j = \{w(s) : s \in S_j\}$. If any vertex w is chosen by more than $2k$ different vertices of S_j then G contains a k-structure of Type 1, a contradiction. Hence if $|S_j| \geqslant 64k^3$, then $|T_j| \geqslant 32k^2$. At least $16k^2$ of these vertices must lie to the left, say, of A_j. Denote these vertices $w_1 < \ldots < w_{16k^2}$, and let $f : [16k^2] \to S_j$ map $x \in [16k^2]$ to a vertex which chose w_x, i.e., $w(f(x)) = w_x$ for each $x \in [16k^2]$. By the Erdős-Szekeres Theorem, there is a subsequence A of this set with length at least $4k$ on which f is monotone. Note that f is injective, so f is in fact strictly monotone on A. If f is increasing on A, then G admits a k-structure of Type 2(a); if it is decreasing then G admits a k-structure of Type 2(b). In either case we contradict one of our assumptions, so $|S_j| < 64k^3$ as claimed.

Now, partition $[n]$ into sets $\{B_1, \ldots, B_{2m+1}\}$ of consecutive vertices, with $m < 128k^4$, as follows. Let $\bigcup_{j=1}^{t} S_j \cup \{i_1, \ldots, i_{t-1}\}$ have elements $a_1 < \ldots < a_m$. Note that $m \leqslant (64k^3 - 1)t + t - 1 \leqslant 128k^4 - 1$ by the comments above, and let $a_0 = 0$ and $a_{m+1} = n + 1$. For each $j \in [m]$, let $B_{2j} = a_j$, and for each $j \in [m + 1]$, let $B_{2j-1} = [a_{j-1} + 1, a_j - 1]$.

Now, each set B_i either consists of a single vertex, or $B_i \subset A_j \setminus S_j$ for some $j \in [t]$, and is an interval of $[n]$. Hence if $x, y \in B_i$, then $\Gamma(x) \setminus A_j = \Gamma(y) \setminus A_j$. Since A_j is ℓ-complete or ℓ-empty, it follows that $x \sim_\ell y$. So we have partitioned $V(G)$ into at most $2m + 1 < 256k^4$ blocks of consecutive vertices, with each block ℓ-homogeneous, as required. □

3 Structures of Type 1, 2 and 3

In order to use Lemma 2.1 to prove Theorem 1.1, we must give sharp lower bounds on the possible speeds of hereditary properties containing large structures of Type 1, 2 or 3. The bounds are given by Lemmas 3.2 and 3.5. Lemma 3.2 will show that if $\mathcal{P}$ contains arbitrarily large structures of Type 1 or 2, then $|\mathcal{P}_n| \geqslant 2^{n-1}$ for every $n \in \mathbb{N}$. To prove it we will need to handle some particular ordered graphs, and for ease of presentation we first define them here.

Let $m \in \mathbb{N}$, let $X = \{x_1, \ldots, x_m\}$ and $Y = \{y_1, \ldots, y_m\}$ be disjoint sets of vertices satisfying $x_1 < \ldots < x_m < y_1 < \ldots < y_m$, and let $I = (I_1, I_2, I_3, I_4) \in \{0,1\}^4$. The graph $M_I^< = M_I^<(X, Y)$ on $X \cup Y$ has the edge set defined as follows:

(i) $x_i x_j \in E(M_I^<)$ if and only if $I_1 = 1$;
(ii) if $i < j$, then $x_i y_j \in E(M_I^<)$ if and only if $I_2 = 1$;
(iii) if $i > j$, then $x_i y_j \in E(M_I^<)$ if and only if $I_3 = 1$;
(iv) $y_i y_j \in E(M_I^<)$ if and only if $I_4 = 1$;
(v) $x_i y_i \in E(M_I^<)$ if and only if i is odd.

The graph $M_I^> = M_I^>(X, Y)$ is obtained in exactly the same way, except the vertices of Y are first renamed so that $y_1 > \ldots > y_m$.

Lemma 3.1. *Let $n \in \mathbb{N}$, $|X| = |Y| \geqslant n^2 + n$ and $I \in \{0,1\}^4$. Then $M_I^<(X, Y)$ and $M_I^>(X, Y)$ each have at least 2^{n-1} distinct induced ordered subgraphs of order n.*

Proof. Let $I = (I_1, I_2, I_3, I_4)$ and $M = M_I^<(X, Y)$. Assume first, by taking the complementary graph (and removing x_1 and y_1) if necessary, that $I_2 = I_3 = 1$ does not hold, and that either $I_2 = I_3 = 0$ or $I_1 = 0$. In the latter case, we may assume also (by symmetry) that $I_3 = 0$. Note first that the result is clear if $n \leqslant 2$. Now, if $n = 3$ and $I_2 = I_3 = I_4 = 0$, then the four ordered subgraphs induced by the sets $\{x_1, y_1, y_2\}$, $\{x_1, x_2, y_1\}$, $\{x_1, x_3, y_3\}$ and $\{y_1, y_2, y_3\}$ are distinct, and if $n = 3$, $I_2 = I_3 = 0$ and $I_4 = 1$, then the four ordered subgraphs induced by the sets $\{x_1, y_1, y_2\}$, $\{x_1, y_2, y_3\}$, $\{x_1, x_2, y_1\}$ and $\{y_1, y_2, y_3\}$ are distinct, so we are done in these cases as well. Hence we may assume that either $n \geqslant 4$ and $I_2 = I_3 = 0$, or $n \geqslant 3$, $I_2 = 1$ and $I_1 = I_3 = 0$.

We shall describe an injective map ϕ from the subsets of $[n]$ of even size to induced subgraphs of M on n vertices; since there are 2^{n-1} such subsets, this will suffice to prove the lemma. Given a subset $S = \{s_1, \dots, s_\ell, t_1, \dots, t_\ell\} \subset [n]$, with $1 \leqslant s_1 < \dots s_\ell < t_1 < \dots < t_\ell \leqslant n$, we shall define $\phi(S)$ to be a subgraph G of M with a matching, or 'star-matching', between the vertices $\{s_1, \dots, s_\ell\}$ and $\{t_1, \dots, t_\ell\}$ (in a star-matching the edge-set is $\{s_i t_j : i \geqslant j\}$ or $\{s_i t_j : i \leqslant j\}$). This will allow us to reconstruct S from G, and hence show that ϕ is injective.

To be precise, first let

$$A = \{x_i, y_i : i = 2jn - 1, 1 \leqslant j \leqslant \ell\},$$

so $x_i y_i \in E(M)$ for each i with $x_i, y_i \in A$. The vertices of A will correspond to the elements of S. We now need to fill in the space between the vertices of A, but without creating any edges that will prevent us from identifying S (see Figure 1). To this end, let $s_0 = 0$, $s_{\ell+1} = t_1$ and $t_{\ell+1} = n + 1$, and let

$$\begin{aligned} B \;=\; & \{x_i : i \in [(2j-1)n + 1, (2j-1)n + s_j - s_{j-1} - 1] \text{ for some } j \in [\ell + 1]\} \\ \cup \; & \{y_i : i \in [2jn + 1, 2jn + t_{j+1} - t_j - 1] \text{ for some } j \in [\ell]\}. \end{aligned}$$

This is possible because $|X|, |Y| \geqslant n^2 + n - 1$.

Define $\phi(S) = M[A \cup B]$. Notice that this gives $\phi(\emptyset) = E_n$ or K_n, depending on whether $I_1 = 0$ or 1. Also, if $S \neq \emptyset$ then $s_j - s_{j-1} \leqslant n - 1$ for every $j \in [\ell + 1]$, and $t_{j+1} - t_j \leqslant n - 1$ for every $j \in [\ell]$, so $A \cap B = \emptyset$. Therefore

$$\begin{aligned} |A \cup B| \;=\; & 2\ell + \sum_{j=1}^{\ell+1} (s_j - s_{j-1} - 1) + \sum_{j=1}^{\ell} (t_{j+1} - t_j - 1) \\ =\; & 2\ell + (s_{\ell+1} - \ell - 1) + (t_{\ell+1} - t_1 - \ell) \;=\; n. \end{aligned}$$

Moreover, if we identify $A \cup B$ with $[n]$ in the obvious way, then $A = S$. We have two cases to investigate.

Case 1. $I_2 = I_3 = 0$.

Let $G = \phi(S)$ for some even-size subset S of $[n]$, so G is an ordered graph with vertex set $[n]$. We wish to show that S is uniquely determined by G. Let $S = \{s_1, \dots, s_\ell, t_1, \dots, t_\ell\}$ as before, and recall from above that we may assume that $n \geqslant 4$.

If $G = K_n$ then the vertices of G all came from X, or all from Y, since $I_2 = I_3 = 0$ and $n \geqslant 3$. Thus in this case $S = \emptyset$. Also, if $|S| \geqslant 2$ then

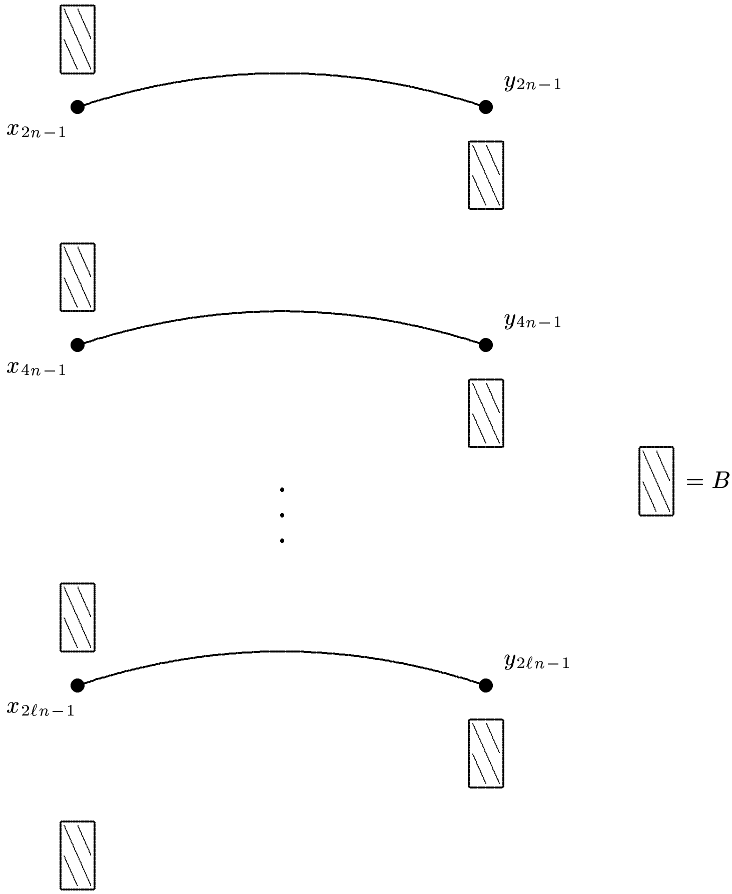


Fig. 1. The set $A \cup B$

$s_1t_1 \in E(G)$, so if $G = E_n$ then we also have $S = \emptyset$ (there is no contradiction here – only one of the two cases is possible for a given M). Therefore we are done in the case $G \in \{K_n, E_n\}$, so assume that G has at least one edge and one non-edge, and hence that $S \neq \emptyset$.

Suppose first that $\Gamma_G(1) \neq \{2\}$, and recall that $G[1, t_1 - 1]$ must be complete or empty. We claim that t_1 must be the left-most vertex of G with a neighbour to its left, but which is not part of a clique involving all of the vertices to its left. To see this, observe that t_1 certainly has a neighbour to its left, since $s_1t_1 \in E(G)$, and that s_1 is its only neighbour to its left, since $I_2 = I_3 = 0$. Hence if t_1 is part of a clique involving all the vertices to its left, then $t_1 = 2$, so $S = \{1, 2\}$ and $\Gamma_G(1) = \{2\}$, a contradiction. Now suppose some vertex to the left of t_1 has a neighbour to its left. Then $I_1 = 1$, so $G[1, t_1 - 1]$ is a clique. This proves the claim, so if $\Gamma_G(1) \neq \{2\}$ then we can reconstruct t_1. But now $S = \{u, v : uv \in E(G) \text{ and } u < t_1 \leqslant v\}$ is the only possibility for S, since $I_2 = I_3 = 0$, so the only edges of $\phi(S)$ between X and Y are the edges $s_i t_i$, for $i \in [\ell]$.

So suppose next that $\Gamma_G(1) = \{2\}$, and suppose also that $\Gamma_G(2) \neq \{1, 3\}$. We shall prove that in this case $S = \{1, 2\}$. Indeed, since $G[1, t_1 - 1]$ must be complete or empty, we have $t_1 \in \{2, 3\}$. Now, if $t_1 = 3$ then $s_1 = 2$, since $s_1 t_1 \in E(G)$ and $13 \notin E(G)$. But then $\Gamma_G(2) = \{1, 3\}$, since $I_2 = I_3 = 0$, which is a contradiction. Hence $t_1 = 2$, and so $S = \{1, 2\}$.

Suppose finally that $\Gamma_G(1) = \{2\}$ and $\Gamma_G(2) = \{1, 3\}$. Since $G[1, t_1 - 1]$ must be complete or empty, we have $t_1 \in \{2, 3\}$, and since $G[t_1, n]$ must also be complete or empty and $n \geqslant 4$, $t_1 = 2$ is impossible. Thus $t_1 = 3$, and so $s_1 \neq 1$, since $3 \notin \Gamma_G(1)$. Hence $S = \{2, 3\}$ in this case.

By the comments above, we have reduced the problem to the following case.

Case 2. $I_1 = 0$, $I_2 = 1$ and $I_3 = 0$.

Recall that we may assume $n \geqslant 3$. Observe that $1t_1 \in E(G)$, so $\Gamma_G(1) = \emptyset$ if and only if $S = \emptyset$. Therefore we may assume that $S \neq \emptyset$.

Consider the homogeneous blocks $B_1, \ldots, B_k$ of G. We claim that they are exactly the sets $\{1, \ldots, s_1\}, \{s_1{+}1, \ldots, s_2\}, \ldots, \{s_\ell{+}1, \ldots, t_1{-}1\}, \{t_1, \ldots, t_2{-}1\}, \ldots, \{t_\ell, \ldots, n\}$. To see this, consider any two vertices $u, v \in G$ and consider the following cases.

(*i*) $s_{j-1} < u < v \leqslant s_j$ for some $j \in [\ell]$. Then $\Gamma(u) \setminus \{v\} = \Gamma(v) \setminus \{u\} = [t_j, n]$, so $u \sim v$.
(*ii*) $s_\ell < u < v < t_1$. Then $\Gamma(u) = \Gamma(v) = \emptyset$, since $I_1 = I_3 = 0$, and u and v are 'below' every vertex of $B \cap Y$. So $u \sim v$.
(*iii*) $t_j \leqslant u < v < t_{j+1}$ for some $j \in [\ell]$ and $I_4 = 0$. Then $\Gamma(u) \setminus \{v\} = \Gamma(v) \setminus \{u\} = [1, s_j]$, so $u \sim v$.
(*iv*) $t_j \leqslant u < v < t_{j+1}$ for some $j \in [\ell]$ and $I_4 = 1$. Then $\Gamma(u) \setminus \{v\} = \Gamma(v) \setminus \{u\} = [1, s_j] \cup [t_1, n] \setminus \{u, v\}$ so $u \sim v$.
(*v*) $u \leqslant s_j < v < t_1$ for some $j \in [\ell]$. Then $ut_j \in E(G)$ but $vt_j \notin E(G)$, so $u \not\sim v$.
(*vi*) $t_1 \leqslant u < t_j \leqslant v$. Then $us_j \notin E(G)$ but $vs_j \in E(G)$, so $u \not\sim v$.
(*vii*) $u = t_1 - 1$ and $v = t_1$. Then $1u \notin E(G)$ and $1v \in E(G)$, so $u \not\sim v$.

Now, there are either 2ℓ or $2\ell + 1$ homogeneous blocks in G (depending on whether or not $t_1 = s_\ell + 1$), and in either case the set S must consist of the right-most vertices of the first ℓ blocks and the left-most vertices of the last ℓ. Thus we have proved that in all cases we can reconstruct S from $\phi(S)$, and hence ϕ is injective. This proves that M has at least 2^{n-1} distinct induced subgraphs, as claimed.

The proof for $M_I^{>}(X, Y)$ is almost identical. □

Lemma 3.1 and Ramsey's Theorem now give the following result.

Lemma 3.2. *Let $\mathcal{P}$ be a hereditary property of ordered graphs. Suppose a k-structure of Type 1 or 2 occurs in $\mathcal{P}$ for arbitrarily large values of k. Then $|\mathcal{P}_n| \geqslant 2^{n-1}$ for every $n \in \mathbb{N}$.*

Proof. Let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose first that $\mathcal{P}$ contains a k-structure of Type 1 for arbitrarily large k. Let $n \in \mathbb{N}$ and choose $k \geqslant n-1$ and an ordered graph $G \in \mathcal{P}$ containing a k-structure of Type 1 on vertices $\{y, x_1, \ldots, x_{2k}\}$. Without loss of generality, assume that $y < x_1 < \ldots < x_{2k}$, and $yx_1 \in E(G)$. Now, for each subset $S \subset [n-1]$, let the ordered graph $G_S \in \mathcal{P}_n$ be induced by the vertices $y \cup \{x_i : i \in (2S-1) \cup (2S^c)\}$, where $2S-1 = \{2s-1 : s \in S\}$ and $2S^c = \{2s : s \in [n-1] \setminus S\}$. The graphs G_S are all distinct, since the set S can be recovered from G_S by considering the neighbours of the left-most vertex, and are all in $\mathcal{P}_n$. It follows that $|\mathcal{P}_n| \geqslant 2^{n-1}$.

So now assume that for some $K \in \mathbb{N}$ there is no k-structure of Type 1 in $\mathcal{P}$ for $k \geqslant K$, and $\mathcal{P}$ contains a k-structure of Type 2 for arbitrarily large k. It can be shown fairly easily that an n-structure of Type 2 contains at least $2^{n/2}$ distinct ordered subgraphs on n vertices. To do better than this we will use Ramsey's Theorem to produce some uniformity on the unknown edges. Let $R_r(s)$ be the smallest number m such that any r-colouring of the edges of K_m contains a monochromatic K_s. Let $n \in \mathbb{N}$, $r = 2^{16}$, $k = R_r(\max\{n^2+n, K+1\})$, and choose an ordered graph $G \in \mathcal{P}$ containing a k-structure of Type 2 on vertices $\{x_1, \ldots, x_{2k}, y_1, \ldots, y_{2k}\}$. We assume without loss of generality that $x_1 < \ldots < x_{2k} < y_i$ for every $i \in [2k]$, that *either* $y_1 < \ldots < y_{2k}$ *or* $y_1 > \ldots > y_{2k}$, and that $x_i y_i \in E(G)$ if and only if i is odd. We shall thus prove the result for both Type 2(a) and Type 2(b) properties at the same time. Let $X = \{x_1, \ldots, x_{2k}\}$ and $Y = \{y_1, \ldots, y_{2k}\}$.

We first split our Type 2 structure up into blocks of four vertices each, as follows: $D_1 = \{x_1, y_1, x_2, y_2\}, D_2 = \{x_3, y_3, x_4, y_4\}, \ldots, D_k = \{x_{2k-1}, y_{2k-1}, x_{2k}, y_{2k}\}$, and let J be the complete graph with these k blocks as vertices. Define a 2^{16}-colouring on the edges of J by associating the bipartite ordered graph $G[D_i, D_j]$ with the edge $D_i D_j$. By our choice of k, there must be a complete monochromatic subgraph of J on $s \geqslant \max\{n^2+n, K+1\}$ blocks. By renaming the vertices of G if necessary, we may assume that these blocks are $D_1, \ldots, D_s$.

Each pair of blocks D_i, D_j (with $1 \leqslant i < j \leqslant s$) induces the same bipartite ordered graph $G[D_i, D_j]$. The next claim shows that there are only a small number of possibilities for this graph.

Claim 3.3. Each of the ordered bipartite graphs $G[\{x_1, x_2\}, \{x_3, x_4\}]$, $G[\{x_1, x_2\}, \{y_3, y_4\}]$, $G[\{x_3, x_4\}, \{y_1, y_2\}]$ and $G[\{y_1, y_2\}, \{y_3, y_4\}]$ is either complete or empty.

Proof. Let $u \in D_1$, and suppose that $\Gamma(u) \cap D_2 \notin \{\emptyset, D_2 \cap X, D_2 \cap Y, D_2\}$. We shall show that $\mathcal{P}$ contains a K-structure of Type 1.

Let $\{v, w\} = D_2 \cap X$ with $v < w$, and suppose that $uv \in E(G)$ and $uw \notin E(G)$ (the proof in the other cases is identical). Consider the set $\{u, x_3, x_4, \ldots, x_{2s}\}$, and note that either $u < x_3$ or $u > x_{2s}$. Now, since $G[D_i, D_j]$ is the same for each $1 \leqslant i < j \leqslant s$, we have $ux_i \in E(G)$ if and only if i is odd. Also, $2s \geqslant 2K + 2$. Thus the graph $G[\{u, x_3, x_4, \ldots, x_{2s}\}] \in \mathcal{P}$ contains a K-structure of Type 1.

We now have a contradiction, so in fact $\Gamma(u) \cap D_2 \in \{\emptyset, D_2 \cap X, D_2 \cap Y, D_2\}$ for every $u \in D_1$. Similarly, one can show that $\Gamma(u) \cap D_1 \in \{\emptyset, D_1 \cap X, D_1 \cap Y, D_1\}$ for every $u \in D_2$. The result now follows easily. □

Now, let $X' = \{x_i : i \in [2s] \text{ and } i \equiv 1, 4 \pmod 4\}$ and $Y' = \{y_i : i \in [2s] \text{ and } i \equiv 1, 4 \pmod 4\}$, and let H be the subgraph of G induced by the vertices of $X' \cup Y'$. Note that no two distinct vertices of X' lie in the same block D_i, and similarly for Y'.

Claim 3.4. For some $I \in \{0,1\}^4$, $H = M_I^{<}(X', Y')$ or $M_I^{>}(X', Y')$.

Proof. It is clear from Claim 3.3, and the fact that $G[D_i, D_j]$ is the same for each $1 \leqslant i < j \leqslant s$, that $H[X']$ and $H[Y']$ are either complete or empty. So let $x_i \in X'$ and $y_j \in Y'$, and observe that

(*i*) if $i < j$ then $x_i y_j \in E(H)$ if and only if $G[\{x_1, x_2\}, \{y_3, y_4\}]$ is complete,
(*ii*) if $i > j$, then $x_i y_j \in E(H)$ if and only if $G[\{x_3, x_4\}, \{y_1, y_2\}]$ is complete,
(*iii*) if $i = j$, then $x_i y_j \in E(H)$ if and only if $i = j$ is odd.

Therefore $H = M_I^{<}(X', Y')$ or $M_I^{>}(X', Y')$ for some $I \in \{0,1\}^4$. □

We have shown that for some $I \in \{0,1\}^4$, either $M_I^{<}(X', Y') \in \mathcal{P}$ or $M_I^{>}(X', Y') \in \mathcal{P}$, with $|X'| = |Y'| \geqslant n^2 + n$. By Lemma 3.1, it follows that $|\mathcal{P}_n| \geq 2^{n-1}$. Since n was arbitrary, the result follows. □

For Type 3 structures, a different bound holds.

Lemma 3.5. *Let $\ell \in \mathbb{N}$, and $\mathcal{P}$ be a hereditary property of ordered graphs. Suppose a (k, ℓ)-structure of Type 3 occurs in $\mathcal{P}$ for arbitrarily large values of k. Then $|\mathcal{P}_n| \geqslant F_{n,\ell+1}$ for every $n \in \mathbb{N}$.*

Proof. Let $\ell \in \mathbb{N}$, and let $\mathcal{P}$ be a hereditary property of ordered graphs containing (k, ℓ)-structures of Type 3 for arbitrarily large values of k. If $\mathcal{P}$ also contains k-structures of Type 1 for arbitrarily large values of k, then by Lemma 3.2, $|\mathcal{P}_n| \geqslant 2^{n-1} \geqslant F_{n,\ell+1}$ for every $n \in \mathbb{N}$, in which case we are done. So assume that there exists $K \in \mathbb{N}$ such that there is no k-structure of Type 1 in $\mathcal{P}$ for any $k \geqslant K$.

Let $n \in \mathbb{N}$, $r = 2^{4(\ell+1)^2}$, $k = R_r(\max\{2n, 2K + 1\})$, and choose a graph $G \in \mathcal{P}$ containing a (k, ℓ)-structure of Type 3. Let the vertices of this Type 3 structure be

$$\{x_i : i \in [2k]\} \cup \{y_i : i \in [2k]\} \cup \{z_{i,j} : i \in [2k], j \in [\ell - 1]\},$$

where $x_1 < z_{1,1} < \ldots < z_{1,\ell-1} < y_1 < \ldots < x_{2k} < z_{2k,1} < \ldots < z_{2k,\ell-1} < y_{2k}$, and without loss of generality $x_i y_i \in E(G)$ if and only if i is odd. We shall apply the same method as in the proof of Lemma 3.2.

As before, group the vertices into blocks, this time of size $2(\ell+1)$, as follows: let $D_i = \{x_{2i-1}, y_{2i-1}, x_{2i}, y_{2i}\} \cup \{z_{i,j} : i \in \{2i-1, 2i\}, j \in [\ell-1]\}$ for $1 \leqslant i \leqslant k$. Note that $D_1 < D_2 < \ldots < D_k$. Let J be the complete graph with these k blocks as vertices, and define a $2^{4(\ell+1)^2}$-colouring on the edges of J by associating the bipartite ordered graph $G[D_i, D_j]$ with the edge $D_i D_j$. By our choice of k, there must be a complete monochromatic subgraph of J on $s \geqslant \max\{2n, K+1\}$ blocks. By renaming the vertices of G if necessary, we may assume that these blocks are $D_1, \ldots, D_s$.

Suppose that some vertex $u \in D_1$ sends an edge and a non-edge to D_2; say $uv_2 \in E(G)$ and $uw_2 \notin E(G)$, with $v_2, w_2 \in D_2$. Since $G[D_i, D_j]$ is the same for every $1 \leqslant i < j \leqslant s$, this means that $uv_j \in E(G)$, and $uw_j \notin E(G)$ for each $j \in [2, s]$, where v_j and w_j are the vertices of D_j corresponding to v_2 and w_2 respectively. Thus u sends an edge and a non-edge (in the same order) to each D_j with $j \in [2, s]$, so $G[\{u, v_2, w_2, \ldots, v_s, w_s\}] \in \mathcal{P}$ contains an $(s-1)$-structure of Type 1. Since $s - 1 \geqslant K$, this is a contradiction, so either $D_2 \subset \Gamma(u)$, or $D_2 \subset \overline{\Gamma(u)}$, for each $u \in D_1$. Similarly, one can show that either $D_1 \subset \Gamma(u)$ or $D_1 \subset \overline{\Gamma(u)}$ for each $u \in D_2$.

It follows easily that the ordered bipartite graph $G[D_1, D_2]$ is complete or empty. Since the ordered graph $G[D_i, D_j]$ is the same for every $i < j$, this implies that either all or none of the edges $\{uv : u \in D_i, v \in D_j, 1 \leqslant i < j \leqslant s\}$ are in $E(G)$. Suppose, by taking the complement of G if necessary, that none of these edges are in $E(G)$, and let H be the subgraph of G induced by the vertices $\{x_i, y_i, z_{i,t} : i \in [s] \text{ is odd, and } t \in [\ell-1]\}$. Note that $x_i y_i \in E(G)$ for each $x_i, y_i \in V(H)$.

We claim that H has at least $F_{m,\ell+1}$ distinct induced ordered subgraphs on m vertices for every $m \leqslant n$. This is clear if $n = 1$, so let $n \geqslant 2$ and suppose the result is true for $n-1$. Then $H[\{x_i, y_i, z_{i,t} : i \in [3, s] \text{ is odd, and } t \in [\ell-1]\}]$ has at least $F_{m,\ell+1}$ distinct subgraphs on m vertices for every $m \leqslant n-1$. It follows that, for $1 \leqslant t \leqslant \ell+1$, H has at least $F_{n-t,\ell+1}$ distinct subgraphs M of order n in which $\max\{v : x_1 v \in E(M)\} = t$ (where $t = 1$ if x_1 is isolated). These are all distinct, so H also has at least $F_{n-1,\ell+1} + F_{n-2,\ell+1} + \ldots + F_{n-(\ell+1),\ell+1} = F_{n,\ell+1}$ distinct subgraphs of order n, as claimed. Since $H \in \mathcal{P}$, and n was arbitrary, the proof of the lemma is complete. □

The following corollary of Lemmas 2.1, 3.2 and 3.5 summarises what we have proved so far.

Corollary 3.6. *Let $\mathcal{P}$ be a hereditary property of ordered graphs. Suppose that $|\mathcal{P}_n| < F_{n,\ell+1}$ for some ℓ and $n \in \mathbb{N}$. Then there exists a $K \in \mathbb{N}$ such that every ordered graph $G \in \mathcal{P}$ may be partitioned into at most K blocks of consecutive vertices, with each block ℓ-homogeneous.*

Proof. Let $\mathcal{P}$ be a hereditary property of ordered graphs, let $\ell, n \in \mathbb{N}$, and suppose that $|\mathcal{P}_n| < F_{n,\ell+1}$. First note that $F_{n,\ell} \leqslant 2^{n-1}$ for every $\ell, n \in \mathbb{N}$, so also $|\mathcal{P}_n| < 2^{n-1}$. Hence, by Lemmas 3.2 and 3.5, there exists $k \in \mathbb{N}$ such that $\mathcal{P}$ contains no k-structure of Type 1 or 2, and no (k, ℓ)-structure of Type 3.

It now follows immediately from Lemma 2.1 that every ordered graph $G \in \mathcal{P}$ may be partitioned into at most $K = 256k^4$ blocks, with each block ℓ-homogeneous. □

4 Polynomial Speed

Before considering the general case, we shall show that if $|\mathcal{P}_n| < F_n$ for some $n \in \mathbb{N}$, then $|\mathcal{P}_n|$ grows only polynomially, and moreover, for sufficiently large n it is exactly a polynomial.

Recall that a set $B \subset V(G)$ is said to be a *homogeneous block* if it is a maximal 1-homogeneous block, i.e., a maximal set of consecutive vertices such that for all $x, y \in B$, $\Gamma(x) \setminus \{y\} = \Gamma(y) \setminus \{x\}$. The *homogeneous block sequence* of G is the sequence $t_1 \geqslant t_2 \geqslant \ldots$, where $t_1, t_2, \ldots$ are the orders of the homogeneous blocks of G. Note that the sequence is uniquely determined, but that $t_1, t_2, \ldots$ is not necessarily the order of the appearance of the blocks.

We need one more piece of notation before we begin. Let G be an ordered graph, and $B_1, \ldots, B_m$ be a collection of 1-homogeneous blocks of G, with $B_1 < \ldots < B_m$ and $V(G) = B_1 \cup \ldots \cup B_m$. (For example, $B_1, \ldots, B_m$ could be the homogeneous blocks of G.) Define $G(B_1, \ldots, B_m)$ to be the ordered graph with possible loops, H, with vertex set $[m]$, and in which $ij \in E(H)$ if and only if $b_i c_j \in E(G)$ for some (and so every) $b_i \in B_i$ and $b_i \neq c_j \in B_j$. Note that a vertex of H has a loop if and only if the corresponding block induces a non-trivial clique (i.e., a clique with at least two vertices).

Let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose that $|\mathcal{P}_n| < F_n$ for some $n \in \mathbb{N}$. By Corollary 3.6, there exists $k \in \mathbb{N} \cup \{0\}$ such that every ordered graph $G \in \mathcal{P}$ has at most $k + 1$ homogeneous blocks. Thus $t_{k+2} = 0$ for every $G \in \mathcal{P}$. The following lemma shows that in this case, the speed is $O(n^k)$.

Lemma 4.1. *Let $\mathcal{P}$ be a hereditary property of ordered graphs, and let $k, M \geqslant 0$ be integers. Suppose that for every $G \in \mathcal{P}$, the homogeneous block sequence of G satisfies* $\displaystyle\sum_{i=k+2}^{\infty} t_i \leqslant M$. *Then $|\mathcal{P}_n| = O(n^k)$.*

Proof. Let $\mathcal{P}$ be a hereditary property of ordered graphs, let $k, M \geqslant 0$ be integers, and suppose that $t_{k+2} + t_{k+3} + \ldots \leqslant M$ for every $G \in \mathcal{P}$. We shall give an upper bound on the number of ordered graphs of order n in the property.

Indeed, every ordered graph $G \in \mathcal{P}_n$ is determined by a sequence $S = (a_1, \ldots, a_m)$ of positive integers, with $1 \leqslant m \leqslant k + M + 1$, $\sum_{i=1}^m a_i = n$, and $\sum_{i \in I} a_i \geqslant n - M$ for some set $I \subset [m]$ with $|I| \leqslant k + 1$; and an ordered graph H, with possible loops, on m vertices. To see this, let $G \in \mathcal{P}_n$ have homogeneous blocks $B_1, \ldots, B_m$ satisfying $B_i < B_j$ if $i < j$, let $a_i = |B_i|$ for each $i \in [m]$, and let $H = G(B_1, \ldots, B_m)$. Now, $1 \leqslant m \leqslant k + M + 1$, since $\sum_{i=k+2}^{\infty} t_i \leqslant M$; $\sum_{i=1}^{m} a_i = n$ since $|G| = n$; and $\sum_{i \in I} a_i \geqslant n - M$ if $I = \{i : B_i$ is one of the largest $k+1$ homogeneous blocks of $G\}$. Thus $S = (a_1, \ldots, a_m)$ and H satisfy the conditions above. It is clear that G can be reconstructed from S and H.

It remains to count the number of such pairs (S, H). If $m \leqslant k+1$ then the number of sequences is just $\binom{n-1}{m-1}$. If $m > k+1$, then a sequence is determined by choosing a subset $I \subset [m]$ of size $k+1$, choosing values $\{a'_i : i \in I\}$ so that $\sum_{i \in I} a'_i = n - M$, and then partitioning $[M]$ into m (possibly empty) intervals $C_1, \ldots, C_m$, and setting $a_i = |C_i|$ if $i \notin I$, and $a_i = a'_i + |C_i|$ if $i \in I$. Thus the number of sequences S is at most

$$\sum_{m=1}^{k+1} \binom{n-1}{m-1} + \sum_{m=k+2}^{k+M+1} \left(\binom{m}{k+1} \binom{n-M+k}{k} \binom{M+m-1}{m-1} \right)$$
$$< \; (k+M+1) \binom{k+M+1}{k+1} 2^{2M+k} \binom{n+k}{k} \; = \; O(n^k).$$

The number of ordered graphs H with possible loops on m vertices is just a constant, so this proves the result. ∎

We next show that in fact, if k is taken to be minimal in Lemma 4.1, then $|\mathcal{P}_n| = \Theta(n^k)$. The following lemma gives the lower bound required to prove this result. If G is an ordered graph, and $u, v, w \in V(G)$, then say that u and v *differ with respect to* w if $uw \in E(G)$ but $vw \notin E(G)$, or vice-versa. This definition can be extended to homogeneous blocks in the obvious way.

Lemma 4.2. *Let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose that there are ordered graphs $G \in \mathcal{P}$ such that t_{k+1}, the size of the $(k+1)^{st}$ largest homogeneous block in G, is arbitrarily large. Then*

$$|\mathcal{P}_n| \geqslant \binom{n-3k-2}{k} = n^k/k! + O(n^{k-1})$$

as $n \to \infty$, and in particular if $k = 1$, then $|\mathcal{P}_n| \geqslant n$ for every $n \in \mathbb{N}$.

Proof. Let $\mathcal{P}$ be a hereditary property of ordered graphs, let $n, k \in \mathbb{N}$, and let $G \in \mathcal{P}$ have $k+1$ homogeneous blocks of order at least n. We shall construct a subgraph H of G with at least $\binom{n-3k-2}{k}$ distinct ordered subgraphs of order n. The idea is simply that H should also have $k+1$ large homogeneous blocks, but at most $2k$ other vertices.

Let $B_1, \ldots, B_{k+1}$ be homogeneous blocks of G, each of order at least n, and with $B_i < B_j$ if $i < j$. Let $V_0 = B_1 \cup \ldots \cup B_{k+1}$ and let $H_0 = G[V_0]$. We shall inductively define a sequence of sets $V_0 \subset V_1 \subset \ldots \subset V_t$, for some $t \in [0, k]$, such that $|V_{i+1}| \leqslant |V_i| + 2$ for each $i \in [1, t-1]$, and so that the sets $\{B_i : i \in [k+1]\}$ are all in different homogeneous blocks of $H = G[V_t]$.

Let $i \in [0, k-1]$, suppose we have already defined the sets $V_0 \subset \ldots \subset V_i$, and let $H_i = G[V_i]$. If the sets $\{B_i : i \in [k+1]\}$ are all in different homogeneous blocks of H_i, then we are done with $t = i$ and $H = H_i$. So suppose that there exists $j \in [k]$ such that B_j and B_{j+1} are in the same homogeneous block of H_i. We shall find a set V_{i+1} as required, such that B_j and B_{j+1} are in different homogeneous blocks of $G[V_{i+1}]$. Note that $G[B_j \cup B_{j+1}]$ must be either complete or empty; without loss of generality, assume that it is empty.

Suppose first that there exists a vertex $u \in V(G) \setminus (B_j \cup B_{j+1})$ such that $B_j \subset \Gamma(u)$ but $B_{j+1} \not\subset \Gamma(u)$, or vice-versa. In this case let $V_{i+1} = V_i \cup \{u\}$. Since B_j and B_{j+1} differ with respect to u, they are in different homogeneous blocks of $G[V_{i+1}]$, as required.

So suppose that every vertex $v \in B_j \cup B_{j+1}$ has exactly the same neighbourhood in G. Since B_j and B_{j+1} are distinct homogeneous blocks of G, this means that there must exist vertices $v, w \in V(G)$ with $B_j < v < B_{j+1}$ and such that v differs from the vertices of $B_j \cup B_{j+1}$ with respect to w. In this case let $V_{i+1} = V_i \cup \{v, w\}$. Again B_j and B_{j+1} are in distinct homogeneous blocks of $G[V_{i+1}]$, as required.

Now, the sequence $(V_0, \ldots, V_t)$ cannot continue any further than $t = k$, since if B_j and B_{j+1} are in different homogeneous blocks of H_i (for some $i \in [0, t-1]$ and $j \in [k]$), then they are in different homogeneous blocks of H_{i+1}. Since each step of the process described above separates B_j and B_{j+1} for at least one $j \in [k]$, after k steps all $k+1$ sets B_i must be in different homogeneous blocks of $H = G[V_t]$.

Now, H has $k+1$ homogeneous blocks of size at least n, and at most $2k$ other vertices, since $|V_{i+1}| \leqslant |V_i| + 2$ for each $i \in [0, t-1]$. Consider an ordered subgraph F of H, which includes all the vertices of $V_t \setminus V_0$, and at least two vertices from each block B_i. The homogeneous blocks of F are $\{V(F) \cap C_i : C_i \text{ is a homogeneous block of } H\}$, and so two such ordered subgraphs with different sequences $(|V(F) \cap B_1|, \ldots, |V(F) \cap B_{k+1}|)$ are distinct. Hence H has at least $\binom{n-3k-2}{k}$ distinct ordered subgraphs (this is the number of sequences of integers $(a_1, \ldots, a_{k+1})$, with $a_i \geqslant 2$ for each $i \in [k+1]$, and $\sum a_i = n - 2k$), and so

$$|\mathcal{P}_n| \geqslant \binom{n-3k-2}{k} = n^k/k! + O(n^{k-1})$$

as required.

To prove the second part of the lemma, let $k = 1$ and perform the same process as above to obtain the ordered graph $H \in \mathcal{P}$. We are left to count the number of subgraphs of H in the various different cases. Let the two large homogeneous blocks be B and C, with $B < C$, and let $n \in \mathbb{N}$. There are four cases to consider.

Case 1. $H = G[B \cup C]$.

H contains either all or none of the edges between B and C; suppose without loss of generality that it contains none. Now, since B and C are distinct homogeneous blocks in H, at least one of B and C must induce a clique. Again without loss, suppose that $H[B]$ is complete.

For each $i \in [n]$, let $H(i)$ denote the ordered subgraph of H which contains i vertices from B, and $n-i$ vertices from C. The leftmost i vertices of $H(i)$ induce a clique, and the leftmost $i+1$ vertices do not. Hence the ordered graphs $\{H(i) : i \in [n]\}$ are all distinct, and are all in $\mathcal{P}_n$. So $|\mathcal{P}_n| \geqslant n$.

Case 2. $H = G[B \cup C \cup \{u\}]$, where $u \notin B \cup C$, $B \subset \Gamma(u)$ and $C \not\subset \Gamma(u)$.

Recall that in this case (and subsequent ones) $G[B \cup C]$ is complete or empty (since u was necessary to distinguish them); assume without loss that it is empty. Now consider the n ordered subgraphs of H obtained by taking i vertices of B, $n-i-1$ vertices of C, and u, for $0 \leqslant i \leqslant n-1$. These subgraphs have exactly i edges, so are different. So $|\mathcal{P}_n| \geqslant n$ in this case too.

Case 3. $H = G[B \cup C \cup \{v\}]$, where $B < v < C$, and $B \cup C \subset \Gamma(v)$.

Assuming again that $G[B \cup C]$ is empty, consider the n ordered subgraphs of H obtained by taking i vertices of B, $n-i-1$ vertices of C, and v, for $0 \leqslant i \leqslant n-1$. They are all distinct if $n \geqslant 3$, since the $(i+1)^{st}$ vertex has degree $n-1$, and all other vertices have degree 1. The result is clear if $n \leqslant 2$, so in this case again $|\mathcal{P}_n| \geqslant n$ (in fact, adding the empty ordered graph, we get $|\mathcal{P}_n| \geqslant n+1$).

Case 4. $H = G[B \cup C \cup \{v, w\}]$, with $B < v < C$, and $E[B \cup C \cup \{v\}] = \emptyset$.

If $vw \in E(H)$, then it is the only edge of H (since v differs from B and C with respect to w), and so the $n-1$ ordered subgraphs of H obtained by taking i vertices of B, $n-i-2$ vertices of C, and the vertices v and w, for $0 \leqslant i \leqslant n-2$, are all distinct. Also none of these ordered graphs is empty, since they all contain the edge vw, so upon adding the empty ordered graph, we get $|\mathcal{P}_n| \geqslant n$.

Similarly, if $vw \notin E(H)$ then $B \cup C \subset \Gamma(w)$, and the same method again gives $|\mathcal{P}_n| \geqslant n$.

Hence $|\mathcal{P}_n| \geqslant n$ in each case. Since n was arbitrary, we are done. □

Remark 4.3. The constant $1/k!$ and the lower bound n in Lemma 4.2 are best possible. To see this, consider the family $\mathcal{P}$ of all ordered graphs with at most k edges, each of length 1, and all independent. It is easy to see that $\mathcal{P}$ is hereditary, and to check that $|\mathcal{P}_n| = \sum_{i=0}^{k} \binom{n-i}{i}$ for every $n \in \mathbb{N}$. We suspect that this is in fact the correct lower bound on speeds of order n^k.

Combining Lemmas 4.1 and 4.2, we obtain the following result.

Corollary 4.4. *Let $\mathcal{P}$ be a hereditary property of ordered graphs. If $|\mathcal{P}_m| < F_m$ for some $m \in \mathbb{N}$, then $|\mathcal{P}_n| = \Theta(n^k)$ for some $k \in \mathbb{N}$, and moreover k is the minimal number such that $\sum_{i=k+2}^{\infty} t_i$ is bounded.*

Proof. By Corollary 3.6 with $\ell = 1$, there exists $K \in \mathbb{N}$ such that every ordered graph $G \in \mathcal{P}$ has at most K homogeneous blocks. Thus $t_j = 0$ for every $j \geqslant K + 1$, so there exists a minimal number k such that $\sum_{i=k+2}^{\infty} t_i$ is bounded. By Lemma 4.1, this implies that $|\mathcal{P}_n| = O(n^k)$. Now since k is minimal, there exist ordered graphs $G \in \mathcal{P}$ such that t_{k+1} is arbitrarily large. Thus, by Lemma 4.2, $|\mathcal{P}_n| = \Omega(n^k)$, so in fact $|\mathcal{P}_n| = \Theta(n^k)$. □

We have proved that $|\mathcal{P}_n| = \Theta(n^k)$ for some $k \in \mathbb{N}$. In fact we can prove a much stronger statement, for which we will need a little preparation. We shall define a set of canonical properties, as in [BBW00], and show that if $|\mathcal{P}_n| = \Theta(n^k)$ with $k \in \mathbb{N}$, then $\mathcal{P}$ is the union of some subset of these properties.

Let $m \in \mathbb{N}$, and suppose that H is an ordered graph with possible loops on $[m]$, and $b : [m] \to \mathbb{N} \cup \{\infty\}$ is any function. Let $\mathcal{P}^*(H, b)$ denote the collection of ordered graphs G which may be partitioned into 1-homogeneous blocks $B_1 < \ldots < B_m$ satisfying $1 \leqslant |B_i| \leqslant b(i)$ for each i, and $G(B_1, \ldots, B_m) = H$. Define $\mathcal{P}(H, b)$ to be the smallest hereditary property of ordered graphs containing $\mathcal{P}^*(H, b)$.

Now, for each ordered graph G with $t_k \neq t_{k+1}$, define the *k-type graph* $H_G^{(k)}$ of G as follows. Let $m = k + n - \sum_{i=1}^{k} t_i$, and partition $[n]$, the vertex set of G, into intervals $B_1 < \ldots < B_m$, so that either B_i is one of the k largest homogeneous blocks of G, so $|B_i| = t_j$ for some $j \in [k]$, or $|B_i| = 1$. Since $t_k \neq t_{k+1}$, the blocks B_i are uniquely determined by G and k; we shall call them the *k-blocks* of G. Let $H_G^{(k)} = G(B_1, \ldots, B_m)$. Thus $H_G^{(k)}$ is uniquely determined by G and k.

Given a set S and a function $b : S \to \mathbb{N} \cup \{\infty\}$, let $I(b) = \{i \in S : b(i) = \infty\}$, and $J(b) = \{i \in S : b(i) > 1\}$. Let $\mathcal{P}$ be a property of ordered graphs, and let $G \in \mathcal{P}$. If G has $t_k \neq t_{k+1}$, then we define the *k-type functions* $\mathcal{B}_G^{(k)}$ of G (with respect to $\mathcal{P}$) as follows. Let $\mathcal{B}_G^{(k)}$ be the set of functions $b : V(H_G^{(k)}) \to \mathbb{N} \cup \{\infty\}$ such that $J(b) = \{i : |B_i| > 1\}$, where $B_1 < \ldots < B_m$ are the k-blocks of G, $\mathcal{P}(H_G^{(k)}, b) \subset \mathcal{P}$, and b is maximal subject to these constraints.

Finally, if $\mathcal{P}(H_G^{(k)}, b) \subset \mathcal{P}$ for every $b : V(H_G^{(k)}) \to \mathbb{N} \cup \{\infty\}$ with $J(b) = \{i : |B_i| > 1\}$, then let $b_G^{(k)}$ denote the unique function $b \in \mathcal{B}_G^{(k)}$. Note that $b_G^{(k)}(i) = \infty$ if $i \in J(b_G^{(k)})$, and $b_G^{(k)}(i) = 1$ otherwise.

Note that if $H_G^{(k)} = H_{G'}^{(k)}$, $b \in \mathcal{B}_G^{(k)}$ and $b' \in \mathcal{B}_{G'}^{(k)}$, then either $b = b'$, or b and b' are incomparable functions. We shall need the following easy observation, which is a well-known result in the theory of well-quasi orderings.

Lemma 4.5. *Let $N \in \mathbb{N}$, and $(b_i)_{i \in \mathbb{N}}$ be a sequence of functions $b_i : [N] \to \mathbb{N} \cup \{\infty\}$. Then the sequence contains a pair of comparable functions. In other words, there exists a pair $i, j \in \mathbb{N}$, with $i \neq j$, such that $b_i(n) \geqslant b_j(n)$ for every $n \in [N]$.*

Proof. First note that, by the pigeonhole principle, we may assume that the functions are everywhere finite. We use induction on N. The statement is trivial for $N = 1$, so let $N \geqslant 2$ and assume all the functions are incomparable. Consider any function, b_1 say. It is not smaller than any other, so there is an index i such that $b_j(i) < b_1(i)$ for infinitely many functions b_j, and hence there is a constant $c < b_1(i)$, such that infinitely many functions have i^{th} co-ordinate c. These functions are incomparable on $[N] \setminus \{i\}$, and we are done by induction. ◻

We are ready to prove the main result of this section, that if $|\mathcal{P}_m| < F_m$ for some $m \in \mathbb{N}$, then $|\mathcal{P}_n|$ is exactly a polynomial for sufficiently large values of n.

Theorem 4.6. *Let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose that $|\mathcal{P}_m| < F_m$ for some $m \in \mathbb{N}$. Then there exist integers $K, N \in \mathbb{N} \cup \{0\}$ and $a_0, \ldots, a_K \in \mathbb{Z}$, such that*

$$|\mathcal{P}_n| = \sum_{i=0}^{K} a_i \binom{n}{i}$$

for every $n \geqslant N$.

Proof. Let $\mathcal{P}$ be a hereditary property of ordered graphs, let $m \in \mathbb{N}$, and suppose $|\mathcal{P}_m| < F_m$. By Corollary 4.4, $|\mathcal{P}_n| = \Theta(n^{k-1})$ as $n \to \infty$ for some $k \in \mathbb{N}$. Moreover, there exists $M \in \mathbb{N}$ such that $\sum_{i=k+1}^{\infty} t_i \leqslant M$ for every $G \in \mathcal{P}$, and there exist ordered graphs $G \in \mathcal{P}$ with arbitrarily large values of t_k.

The proof is by induction on k. Let $k \in \mathbb{N}$, and assume the result holds for all smaller values of k. We begin by removing those ordered graphs in $\mathcal{P}$ for which $t_k = t_{k+1}$ is possible. Let

$$\mathcal{P}^{(2)} = \{G \in \mathcal{P} : \sum_{i=k}^{\infty} t_i \leqslant 2M\},$$

and observe that $\mathcal{P}^{(2)}$ is hereditary. Thus $|\mathcal{P}_n^{(2)}| = O(n^{k-2})$, by Corollary 4.4 (if $k = 1$ then $|\mathcal{P}_n^{(2)}| = 0$ for $n \geqslant 2M + 1$). Observe also that $t_k \geqslant M + 1 > t_{k+1}$ for every $G \in \mathcal{P} \setminus \mathcal{P}^{(2)}$.

Next, we shall remove those ordered graphs $G \in \mathcal{P} \setminus \mathcal{P}^{(2)}$ for which $b_G^{(k)}$ is not defined, i.e., for which $|\mathcal{B}_G^{(k)}| \geqslant 2$. (Here, and throughout the proof, $\mathcal{B}_G^{(k)}$ and $b_G^{(k)}$ are taken with respect to $\mathcal{P}$.) Note that since $t_k \neq t_{k+1}$ for every $G \in \mathcal{P} \setminus \mathcal{P}^{(2)}$, $\mathcal{B}_G^{(k)}$ is defined for these G. Let $\mathcal{A} = \{G \in \mathcal{P} \setminus \mathcal{P}^{(2)} : |\mathcal{B}_G^{(k)})| \geqslant 2\}$, and let

$$\mathcal{P}^{(3)} = \bigcup_{G \in \mathcal{A},\, b \in \mathcal{B}_G^{(k)}} \mathcal{P}(H_G^{(k)}, b).$$

Now, $\mathcal{P}^{(3)}$ is hereditary, since $\mathcal{P}(H, b)$ is hereditary for every H and b. We claim that $|\mathcal{P}_n^{(3)}| = O(n^{k-2})$, i.e., that t_k is bounded in $\mathcal{P}^{(3)}$.

In order to prove the claim, let $\mathcal{H} = \{H_G^{(k)} : G \in \mathcal{P} \backslash \mathcal{P}^{(2)}\}$ and observe that each $H \in \mathcal{H}$ has at most $k + M$ vertices, since $|V(H_G^{(k)})| = k + n - \sum_{i=1}^k t_i \leqslant k+M$ for every $G \in \mathcal{P} \backslash \mathcal{P}^{(2)}$. Thus $|\mathcal{H}|$ is finite. Now recall that for any $H \in \mathcal{H}$, and any pair $G, G' \in \mathcal{P} \setminus \mathcal{P}^{(2)}$, if $H_G^{(k)} = H_{G'}^{(k)} = H$, $b \in \mathcal{B}_G^{(k)}$ and $b' \in \mathcal{B}_{G'}^{(k)}$, then either $b = b'$, or b and b' are incomparable. Thus, by Lemma 4.5, there are only finitely many such functions b for each $H \in \mathcal{H}$, and so there are only finitely many pairs (H, b) such that $H = H_G^{(k)}$ and $b \in \mathcal{B}_G^{(k)}$ for some $G \in \mathcal{A}$.

Let $\mathcal{C} = \{(H, b) : H = H_G^{(k)}$ and $b \in \mathcal{B}_G^{(k)}$ for some $G \in \mathcal{A}\}$, and observe that if $(H, b) \in \mathcal{C}$, then $|I(b)| < k$. Therefore, there is an $N(b) \in \mathbb{N}$ such that $t_k \leqslant N(b)$ for every $G \in \mathcal{P}(H, b)$. Since $\mathcal{C}$ is finite, it follows that there exists an $N \in \mathbb{N}$ such that $t_k \leqslant N$ for every $G \in \bigcup_{(H,b) \in \mathcal{C}} \mathcal{P}(H, b) = \mathcal{P}^{(3)}$, as claimed. We choose such an N, with $N \geqslant 2M + 1$. We have $\sum_{i=k}^{\infty} t_i \leqslant N + M$ for every $G \in \mathcal{P}^{(3)}$, so let

$$\mathcal{P}^{(4)} = \{G \in \mathcal{P} : \sum_{i=k}^{\infty} t_i \leqslant N + M\},$$

and observe that $\mathcal{P}^{(4)}$ is hereditary, and that $\mathcal{P}^{(2)} \cup \mathcal{P}^{(3)} \subset \mathcal{P}^{(4)}$. By Lemma 4.1, $|\mathcal{P}_n^{(4)}| = O(n^{k-2})$.

We shall apply the induction hypothesis to the property $\mathcal{P}^{(4)}$, but first let us count the members of $(\mathcal{P} \setminus \mathcal{P}^{(4)})_n$. Let $\mathcal{P}^{(1)} = \mathcal{P} \setminus \mathcal{P}^{(4)}$, and consider an ordered graph $G \in \mathcal{P}^{(1)}$. Since $\mathcal{P}^{(2)} \cup \mathcal{P}^{(3)} \subset \mathcal{P}^{(4)}$, we know that $t_k > t_{k+1}$, and that $b_G^{(k)}$ is defined. Recall that $|I(b_G^{(k)})| = k$. Let $\mathcal{D} = \{(H, b) : H = H_G^{(k)}$ and $b = b_G^{(k)}$ for some $G \in \mathcal{P}^{(1)}\}$.

We claim that $G \in \mathcal{P}^*(H, b)$ for a unique pair $(H, b) \in \mathcal{D}$. Clearly $G \in \mathcal{P}^*(H_G^{(k)}, b_G^{(k)})$, so suppose that also $G \in \mathcal{P}^*(H', b')$, with $(H', b') \in \mathcal{D}$. Then $H' = H_{G'}^{(k)}$ and $b' = b_{G'}^{(k)}$ for some $G' \in \mathcal{P}^{(1)}$, and also $G(B_1, \ldots, B_m) = G'(B_1', \ldots, B_m') = H'$, where $B_1 < \ldots < B_m$ are 1-homogeneous blocks of G, with $V(G) = B_1 \cup \ldots \cup B_m$ and $1 \leqslant |B_i| \leqslant b'(i)$ for each $i \in [m]$, and $B_1' < \ldots < B_m'$ are the k-type blocks of G'. Note that $|I(b')| = |J(b')| = k$.

Now, if $i \in I(b')$, then B_i' is a homogeneous block of G'. Furthermore, if $i \in I(b')$ then $|B_i'| \geqslant N + 1 > 1$, since $G' \in \mathcal{P}^{(1)}$, and if $i \notin I(b') = J(b')$, then $|B_i| = |B_i'| = 1$. So, if $|B_i| > 1$ for each $i \in I(b')$, then (since B_i' is a homogeneous block) it follows that B_i must be a homogeneous block of G for each $i \in I(b')$. Recall that $|H'| = m \leqslant k + M$. We shall show that $|B_i| \geqslant M + 1$ for every $i \in I(b')$, so that in fact, $\{B_i : i \in I(b')\}$ are the largest k homogeneous blocks of G.

Indeed, suppose that $|B_i| \leqslant M$ for some $i \in I(b')$. Note that since the sets B_i are 1-homogeneous blocks, each is contained in some homogeneous block of G. Since $|B_i| = 1$ if $i \notin I(b')$, it follows that $t_k \leqslant |B_i| + (|H| - k) \leqslant 2M$. But $G \in \mathcal{P}^{(1)}$, which implies that $t_k \geqslant N \geqslant 2M + 1$, so this is a contradiction. Hence $|B_i| \geqslant M + 1$ for every $i \in I(b')$, as claimed.

Since $|B_i| > 1$ for every $i \in I(b')$, B_i is a homogeneous block for every $i \in I(b')$. Thus, since $|H'| \leqslant k + M$ and $|B_i| \geqslant M + 1$ for every $i \in I(b')$,

$\{B_i : i \in I(b')\}$ are the largest k homogeneous blocks of G, which implies that $H' = H_G^{(k)}$, by the definition of $H_G^{(k)}$. We have also shown that $b_G^{(k)}(i) = b'(i) = \infty$ if $i \in I(b')$, and 1 otherwise, and so $(H', b') = (H_G^{(k)}, b_G^{(k)})$. Since (H', b') were arbitrary, this shows that $(H_G^{(k)}, b_G^{(k)})$ is the unique pair (H, b) such that $G \in \mathcal{P}^*(H, b)$. Call $(H_G^{(k)}, b_G^{(k)})$ the pair 'realised by' G.

It remains to count how many ordered graphs $G \in \mathcal{P}^{(1)}$ of order n realise a given pair $(H, b) \in \mathcal{D}$. Let $n > k(N+2M+1)+M$. Each vertex $i \in V(H) \setminus J(b)$ is assigned one vertex of G, and each of the remaining k vertices of H must be assigned at least $N + M + 1 - (|V(H)| - k)$ vertices of G, by the definition of $\mathcal{P}^{(1)}$. The remaining $n' = n - k(N + M + 1 - (|V(H)| - k)) - (|V(H)| - k)$ vertices of G may then be assigned arbitrarily to the vertices of $J(b)$. Hence there are exactly $\binom{n'+k-1}{k-1}$ ordered graphs in $\mathcal{P}_n^{(1)}$ which realise a given pair $(H, b) \in \mathcal{D}$. Therefore

$$|\mathcal{P}_n^{(1)}| = \sum_{(H,b) \in \mathcal{D}} \binom{n - K(H,b)}{k-1},$$

where $K(H, b) = k(N + M + k - 1) - (k - 1)|V(H)| - 1$.

Now, when $k = 1$ we have shown that $|\mathcal{P}_n| = |\mathcal{D}|$ for sufficiently large values of n, so the lemma holds in the base case. Let $k \geqslant 2$, and assume that result holds for all smaller values of k. Since $\mathcal{P}^{(4)}$ is a hereditary property of ordered graphs with speed $\Theta(n^{k-2})$, it has speed equal to some polynomial for sufficiently large n, by the induction hypothesis. Also, $|\mathcal{P}_n^{(1)}| = \sum_{(H,b) \in \mathcal{D}} \binom{n-K(H,b)}{k-1}$ for sufficiently large n, and $\binom{n-K}{k-1} = \binom{n}{k-1} - \sum_{i=1}^{K} \binom{n-i}{k-2}$ for any $K \in \mathbb{N}$, so $|\mathcal{P}_n^{(1)}|$ is exactly a polynomial (i.e., it may be written in the form stated in the theorem). Since $\mathcal{P}_n = \mathcal{P}_n^{(1)} \cup \mathcal{P}_n^{(4)}$ and these sets are disjoint, the induction step follows, and hence so does the theorem. □

5 ℓ-empty Ordered Graphs

Let us now return to the general case. We know that $|\mathcal{P}_n| < 2^{n-1}$ for some $n \in \mathbb{N}$, so $|\mathcal{P}_n| < F_{n,\ell+1}$ for some $\ell \in \mathbb{N}$, since $F_{n,\ell+1} = 2^{n-1}$ if $\ell \geqslant n - 1$. Hence, by Corollary 3.6, there exist integers k and ℓ such that every ordered graph $G \in \mathcal{P}$ may be partitioned into at most k blocks of consecutive vertices, with each block ℓ-homogeneous. Before we can deduce the speed of such a property, we need to know more about the ordered graphs induced by these ℓ-homogeneous blocks. The lemmas in this section will allow us to describe them quite precisely. We start with some definitions, which will make our results much easier to state.

Let G be an ordered graph and $u, v \in V(G)$ (with $u < v$). We say that the pair u, v *separates the edges* of G if for every edge $ij \in E(G)$ with $i < j$, either $j \leqslant u$ or $v \leqslant i$. We will call an ordered graph G *irreducible* if no pair of vertices separate the edges of G.

Given any ordered graph G, the vertex set of G can be decomposed in a unique way into blocks of consecutive vertices such that each block induces an irreducible subgraph, and there are no edges between different blocks. We call this the *irreducible block decomposition* of G, and write $B(G) = (G_1, \ldots, G_m)$ if the irreducible blocks of G induce the ordered graphs $G_1, \ldots, G_m$ in that order.

To be more precise, let $G_1, \ldots, G_m$ be (not necessarily distinct) ordered graphs, let n_i be the order of G_i for each $i \in [m]$, and let $N_i = \sum_{j=1}^{i} n_j$ for each $i \in [0, m]$. Define $G_1 + \ldots + G_m$ to be the ordered graph with $V(G_1 + \ldots + G_m) = [N_m] = n_1 + \ldots + n_m$, and

$$E(G_1 + \ldots + G_m) = \bigcup_{i=1}^{m} \{uv : u, v \in [N_{i-1} + 1, N_i] \\ \text{and } (u - N_{i-1})(v - N_{i-1}) \in E(G_i)\}.$$

Observe that an ordered graph G has the irreducible block decomposition $B(G) = (G_1, \ldots, G_m)$ if and only if $G = G_1 + \ldots + G_m$, and each G_i is irreducible. If $B(G) = (G_1, \ldots, G_m)$ and $|G_i| = n_i$ for each $i \in [m]$, then write $BS(G) = (n_1, \ldots, n_m)$ and call this the *irreducible block sequence* of G. Finally, if G is an ordered graph, define $S_n(G)$ to be the number of distinct (i.e., non-isomorphic) induced ordered subgraphs of G of order n.

We begin with a simple observation.

Observation 5.1. *Let $n, k \in \mathbb{N}$ with $k \leqslant n$. If G is an irreducible ordered graph of order n, then there exists an irreducible ordered subgraph of G of order k.*

Proof. The proof is by induction on n. For $n = 1$ and for $k = n$ the result is trivial, so let $n \geqslant 2$ and $k \leqslant n - 1$. Let u be maximal so that $1u \in E(G)$ and v be maximal so that $2v \in E(G)$. If $u \geqslant v$ (or v does not exist) then remove vertex 2; if $u < v$ then remove vertex 1. The resulting ordered graph is clearly irreducible, and has order $n - 1$, so we are done by induction. ☐

Our first lemma controls the number of irreducible blocks of size at least ℓ.

Lemma 5.2. *Let G be an ordered graph, and let $k, \ell \in \mathbb{N}$. If there are at least k blocks of order at least ℓ in the irreducible block decomposition of G, then $S_n(G) \geqslant F_{n,\ell}$ for each $n \leqslant k$.*

Proof. Let $n, k \in \mathbb{N}$ with $n \leqslant k$, and let $G' = G_1 + \ldots + G_k$ be a subgraph of G induced by k of the irreducible blocks of G which are of size at least ℓ. Thus G_i is irreducible for each $i \in [k]$. For each sequence $(a_1, \ldots, a_t)$ with $t \in \mathbb{N}$, $a_i \in [\ell]$ for each $i \in [t]$ and $\sum_i a_i = n$, choose a subgraph of G' with irreducible block sequence $(a_1, \ldots, a_t)$; such a subgraph exists by Observation 5.1, and because $n \leqslant k$.

These subgraphs are all distinct (since they have distinct irreducible block sequences), so it only remains to count them. It is easy to see that there are exactly $F_{n,\ell}$ sequences $(a_1, \ldots, a_t)$ as described above. Therefore G has at least this many distinct subgraphs of order n. □

Using Lemma 5.2, we can now control the size of the largest irreducible block.

Lemma 5.3. *Let $k, \ell \in \mathbb{N}$, and let G be an ℓ-empty ordered graph. If there is a block of size at least $4k\ell$ in the irreducible blocks decomposition of G, then $S_n(G) \geqslant F_{n,\ell}$ for each $n \leqslant k$.*

Proof. Let $k, \ell \in \mathbb{N}$ with $n \leqslant k$, and let G be an ℓ-empty ordered graph with an irreducible block B of size $m \geqslant 4k\ell$. Let G' be the irreducible subgraph of G induced by B, with vertex set $[m]$. We shall find a subgraph H of G' for which $B(H)$ contains k irreducible blocks of size at least ℓ.

Since G' is irreducible, for each vertex $j \in [m]$ there exists at least one edge $uv \in E(G')$ with $j \in [u, v]$. Thus, for each $i \in [k]$, we may define

$$u_i = \min\{u : \exists v, uv \in E(G') \text{ and } (4i-3)\ell \in [u, v]\},$$

and

$$v_i = \max\{v : \exists u, uv \in E(G') \text{ and } (4i-2)\ell \in [u, v]\}.$$

Consider the set of vertices $\{u_1, v_1, \ldots, u_k, v_k\}$. For every $i \in [k]$, $u_i \leqslant (4i-3)\ell$ and $v_i \geqslant (4i-2)\ell$, so $v_i - u_i \geqslant \ell$. Also, since G is ℓ-empty, $u_i \geqslant (4i-4)\ell + 1$ and $v_i \leqslant (4i-1)\ell - 1$ for every $i \in [k]$. Hence $v_i + \ell < u_{i+1}$ for every $i \in [k-1]$.

Set $A = [u_1, v_1] \cup \ldots \cup [u_k, v_k]$, and let $H = G'[A]$. We claim that $B(H) = (G'[u_1, v_1], \ldots, G'[u_k, v_k])$. Since $v_i - u_i \geqslant \ell$ for each $i \in [k]$, this implies that H has k irreducible blocks of size at least ℓ, so, by Lemma 5.2, it will suffice to prove the lemma.

We must show that $H_i = G'[u_i, v_i]$ is irreducible for each $i \in [k]$, and that there are no edges in G' between $[u_i, v_i]$ and $[u_j, v_j]$ if $i \neq j$. The latter statement follows because G is ℓ-empty, and $v_i + \ell < u_{i+1}$ for every $i \in [k-1]$, as observed above. To establish the former statement, suppose that H_i is not irreducible for some $i \in [k]$. Then there must be some consecutive pair $x, y \in [u_i, v_i]$ which separates the edges of H_i. By the definitions of u_i and v_i, there exist edges $u_i v$ and uv_i in G' with $v \geqslant (4i-3)\ell$ and $u \leqslant (4i-2)\ell$, so $x, y \in [v, u] \subset [(4i-3)\ell, (4i-2)\ell]$. Since G' is irreducible, there exists an edge $ab \in E(G')$ with $a \leqslant x, y \leqslant b$, and since x and y separate the edges of H_i, we must have either $a < u_i$ or $b > v_i$. In either case we have a contradiction, since u_i was chosen to be minimal, and v_i was chosen to be maximal. This contradiction proves that H_i is irreducible, so $B(H) = (G'[u_1, v_1], \ldots, G'[u_k, v_k])$, as claimed. □

The next lemma will allow us to tell which graphs may be induced by arbitrarily many irreducible blocks.

Lemma 5.4. *Let $k \in \mathbb{N}$, let $G_1, \ldots, G_k$ be ordered graphs, and let $G = G_1 + \ldots + G_k$. Suppose that for each $i \in [k]$ there is an integer $m(i)$ such that G_i has at least two irreducible induced subgraphs on $m(i)$ vertices. Then $S_n(G) \geqslant 2^{n-1}$ for each $n \leqslant k$.*

Proof. We prove the lemma by induction on k. For $k = 1$ it is trivial, so let $k \geqslant 2$ and assume it is true for $k-1$. Let $G = G_1 + \ldots + G_k$ be an ordered graph as described, and let $G' = G_2 + \ldots + G_k$. By the induction hypothesis applied to G', we have

$$S_n(G) \geqslant S_n(G') \geqslant 2^{n-1}$$

for every $n \leqslant k-1$. We must show that $S_k(G) \geqslant 2^{k-1}$.

Let $m = m(1)$, so G_1 has two irreducible subgraphs of order m, H_m and H'_m. By Observation 5.1, G_1 also has an irreducible subgraph H_j on j vertices for each $j \leqslant m-1$.

Now, by the induction hypothesis, G has at least 2^{k-j-1} induced subgraphs of order k whose left-most irreducible block is H_j, for each $1 \leqslant j \leqslant m-1$. Also G has at least 2^{k-m-1} induced subgraphs (of order k) whose left-most irreducible block is H_m, and at least 2^{k-m-1} whose left-most irreducible block is H'_m. These induced subgraphs are all distinct (since they have different left-most irreducible blocks). Therefore,

$$S_k(G) \geqslant \sum_{j=1}^{m-1} 2^{k-j-1} + 2 \cdot 2^{k-m-1} = 2^{k-1}.$$

This proves the induction step, and so also the lemma. □

We now define the following five collections of (irreducible) ordered graphs, and two sporadic examples, which will play a central role in our characterisation of properties with speed $p(n)F_{n,k}$. Our reasons for choosing these particular ordered graphs will be made clear by Lemma 5.5. For each $n \in \mathbb{N}$, let

- $J_1^{(n)} = K_n$,
- $J_2^{(n)}$ have vertex set $[n]$ and edge set $E = \{1n\}$ (if $n \geqslant 2$),
- $J_3^{(n)}$ have vertex set $[n]$ and edge set $E = \{1i : i \in [2,n]\}$,
- $J_4^{(n)}$ have vertex set $[n]$ and edge set $E = \{in : i \in [n-1]\}$,
- $L^{(n)}$ have vertex set $[n]$ and edge set $E = \{i(i+1) : i \in [n-1]\}$,
- Q_1 have vertex set $\{1,2,3,4\}$ and edge set $\{13, 24\}$,
- Q_2 have vertex set $\{1,2,3,4\}$ and edge set $\{14, 23\}$.

Also let $\mathcal{J}_k = \{J_i^{(n)} : i \in [4], n \leqslant k\} \cup \{L^{(n)} : n \leqslant k\}$ for $k = 1,2,3$ and let $\mathcal{J}_k = \{J_i^{(n)} : i \in [4], n \leqslant k\} \cup \{L^{(n)} : n \leqslant k\} \cup \{Q_1, Q_2\}$ for each $k \geqslant 4$. Finally, let $\mathcal{J} = \bigcup_{k \in \mathbb{N}} \mathcal{J}_k$.

Lemma 5.5. *Let G be a finite irreducible ordered graph, and suppose that G has at most one irreducible ordered subgraph of order* 3, *and at most one of order* 4. *Then $G \in \mathcal{J}$.*

Proof. The result is proved by a simple case analysis, as follows. Let G be an irreducible ordered graph with vertex set $[n]$, and let $t = \max\{|i-j| : ij \in E(G)\}$ be the length of the longest edge in G. Suppose that G has at most one irreducible ordered subgraph of order i for $i = 3, 4$, and that $G \notin \mathcal{J}$. If $t = 1$ then $G = L^{(n)} \in \mathcal{J}$, since G is irreducible, so $t \geqslant 2$.

Let $i(i+t) \in E(G)$ be an edge of maximal length in G, and suppose first that $i+t < n$. Since G is irreducible, the pair $(i+t, i+t+1)$ does not separate the edges of G, so there must be an edge $uv \in E(G)$ with $i < u \leqslant i+t < v$. Now, if $u < i+t$ and $t \geqslant 3$ then the subgraphs $G[\{i, i+1, i+2, i+t\}]$ and $G[\{i, u, i+t, v\}]$ are distinct (since $iv \notin E(G)$), irreducible subgraphs of G, each on 4 vertices, a contradiction. If $u = i+t$ then $G[\{i, i+t, v\}]$ and $G[\{i, i+1, i+t\}]$ are distinct (again since $iv \notin E(G)$), irreducible subgraphs of G, each on 3 vertices, another contradiction. So $t = 2$ and $u = i+1$, i.e., $i(i+2)$ and $(i+1)(i+3) \in E(G)$.

Now, if $j(j+1) \in E(G)$ for some $j \in \{i, i+1, i+2\}$, then $G[i, i+3]$ has at least two irreducible subgraphs on 3 vertices, a contradiction. So if $n = 4$ then $G = Q_1 \in \mathcal{J}$, another contradiction, so $n \geqslant 5$. Now either $i > 1$ or $i+3 < n$; assume without loss (by symmetry) that $i+3 < n$. Now, applying to $(i+1)(i+3)$ the same argument that we used for the edge $i(i+t)$, and using the fact that $t = 2$, we obtain $(i+2)(i+4) \in E(G)$. But now $G[\{i, i+1, i+2\}]$ and $G[\{i, i+2, i+4\}]$ are distinct (since $i(i+4) \notin E(G)$) and irreducible subgraphs of G, each on 3 vertices, a final contradiction. Therefore $i+t = n$, and similarly it can be proved that $i = 1$.

We have shown that $1n \in E(G)$. Suppose that for some pair $1 < i, j < n$ we have $1i \in E(G)$ and $1j \notin E(G)$. Then $G[\{1, i, n\}]$ and $G[\{1, j, n\}]$ are distinct and irreducible subgraphs of G, each on 3 vertices, a contradiction. So $\Gamma(1) = \{n\}$ or $[2, n]$, and similarly $\Gamma(n) = \{1\}$ or $[n-1]$.

Suppose first that $\Gamma(1) = \{n\}$ and $\Gamma(n) = \{1\}$. If $G[2, n-1] \notin \{E_{n-2}, K_{n-2}\}$ then it has an edge ij and a non-edge uv, and $G[\{1, i, j, n\}]$ and $G[\{1, u, v, n\}]$ are distinct and irreducible. If $G[2, n-1] = K_{n-2}$ then either $G = Q_2 \in \mathcal{J}$ (if $n = 4$), or $G[\{1, 2, n\}]$ and $G[\{2, 3, 4\}]$ are distinct and irreducible (if $n \geqslant 5$). If $G[2, n-1] = E_{n-2}$ then $G = J_2^{(n)} \in \mathcal{J}$. In each case we have a contradiction.

The remaining cases are now easy to deal with. If $\Gamma(1) = [2, n]$ and $\Gamma(n) = \{1\}$ then either $G[2, n-1]$ contains an edge ij, in which case $G[\{1, i, j\}]$ and $G[\{1, i, n\}]$ are distinct and irreducible, or $G[2, n-1] = E_{n-2}$, in which case $G = J_3^{(n)} \in \mathcal{J}$. Similarly if $\Gamma(1) = \{n\}$ and $\Gamma(n) = [n-1]$ then either $G[\{i, j, n\}]$ and $G[\{1, i, n\}]$ are distinct and irreducible or $G = J_4^{(n)} \in \mathcal{J}$. Finally, if $\Gamma(1) = [2, n]$ and $\Gamma(n) = [n-1]$ then either there is a non-edge ij in $G[2, n-1]$, in which case $G[\{1, i, j\}]$ and $G[\{i, j, n\}]$ are distinct and irreducible, or $G[2, n-1] = K_{n-2}$, in which case $G = J_1^{(n)} \in \mathcal{J}$. In each case we have a contradiction, so the assumed ordered graph G is impossible, and the proof is complete. □

Combining Lemmas 5.4 and 5.5, we obtain the following result.

Lemma 5.6. *Let $k, m \in \mathbb{N}$, and let G be an ordered graph with irreducible block decomposition $B(G) = (G_1, \ldots, G_m)$. If $|\{i \in [m] : G_i \notin \mathcal{J}\}| \geqslant k$, then $S_n(G) \geqslant 2^{n-1}$ for every $n \leqslant k$.*

Proof. Let $k, m \in \mathbb{N}$, and let G be an ordered graph whose irreducible block decomposition $B(G) = (G_1, \ldots, G_m)$ satisfies $|\{i \in [m] : G_i \notin \mathcal{J}\}| \geqslant k$. Let $\{a(1), \ldots, a(k)\} \subset \{i \in [m] : G_i \notin \mathcal{J}\}$, and let $G' = G_{a(1)} + \ldots + G_{a(k)} \leqslant G$.

Now, for each $i \in [k]$ we have $G_{a(i)} \notin \mathcal{J}$, so by Lemma 5.5 there is an integer $m(i)$ such that $G_{a(i)}$ has at least two irreducible induced subgraphs on $m(i)$ vertices (and in fact it can be assumed that $m(i) \in \{3, 4\}$). Thus, by Lemma 5.4, we have $S_n(G) \geqslant S_n(G') \geqslant 2^{n-1}$ for every $n \leqslant k$. □

Given an ordered graph G and $k \in \mathbb{N}$, define $G^k = G + \ldots + G$, where G appears k times in the sum.

Lemma 5.7. *Let $k \in \mathbb{N}$, let $A, B \in \mathcal{J}$ with $A \nleqslant B$ and $B \nleqslant A$, and let $G = (A + B)^k$. Then $S_n(G) \geqslant 2^{n-1}$ for every $n \leqslant k$.*

Proof. Note that if $A, B \in \mathcal{J}$, with $A \nleqslant B$ and $B \nleqslant A$, then $A + B$ has at least two distinct irreducible ordered subgraphs on $\min(|A|, |B|)$ vertices. Setting $G_i = A + B$ for $1 \leqslant i \leqslant k$, the result follows now immediately by Lemma 5.4. □

For each $k, \ell \in \mathbb{N}$, let $\mathcal{J}(k, \ell)$ be the following collection of ordered graphs. Let $G \in \mathcal{J}(k, \ell)$ if and only if there exist an (ordered) collection of $s \leqslant k$ ordered graphs $(A_1, \ldots, A_s)$ satisfying the following conditions.

- $G = A_1 + \ldots + A_s$, and
- For each $i \in [s]$, there exists an (ordered) collection of ordered graphs $(B_1^{(i)}, \ldots, B_{t(i)}^{(i)})$, satisfying
  - $A_i = B_1^{(i)} + \ldots + B_{t(i)}^{(i)}$,
  - $B_j^{(i)} \in \mathcal{J}_\ell$ for each $j \in [t(i)]$, and
  - $B_j^{(i)} \leqslant B_{j'}^{(i)}$ or $B_{j'}^{(i)} \leqslant B_j^{(i)}$ for each pair $j, j' \in [t(i)]$.

We call a collection $(A_1, \ldots, A_s)$ satisfying these conditions a (k, ℓ)-*witness set* for G.

Lemma 5.8. *Let $k, \ell, m \in \mathbb{N}$ with $k \geqslant 50m\ell^2 + 1$, let G be an ordered graph, and suppose that $G \in \mathcal{J}(k, \ell)$ but $G \notin \mathcal{J}(k-1, \ell)$. Then $S_n(G) \geqslant 2^{n-1}$ for every $n \leqslant m$.*

Proof. Let $k, \ell, m \in \mathbb{N}$, with $k \geqslant 50m\ell^2 + 1$, let G be an ordered graph, and suppose that $G \in \mathcal{J}(k,\ell)$ but $G \notin \mathcal{J}(k-1,\ell)$. Let $\{A_1, \ldots, A_k\}$ be a (k,ℓ)-witness set for G. Since k was chosen to be minimal, for each $i \in [k-1]$ there must exist irreducible blocks $D_i \subset A_i$ and $D_i' \subset A_{i+1}$ such that $G[D_i] \nleqslant G[D_i']$ and $G[D_i'] \nleqslant G[D_i]$; otherwise for some i the collection $(A_1, \ldots, A_{i-1}, A_i + A_{i+1}, A_{i+2}, \ldots, A_k)$ would be a $(k-1,\ell)$-witness set for G.

Consider the multiset of pairs $\mathcal{D} = \{(D_{2i-1}, D'_{2i-1}) : 2i \leqslant k\}$. Since $k \geqslant 50m\ell^2+1$, $|\mathcal{D}| \geqslant 25m\ell^2$. Now, note that there are fewer than 5ℓ ordered graphs in $\mathcal{J}_\ell$, and that D_{2i-1} and $D'_{2i-1} \in \mathcal{J}_\ell$ for each i with $2i-1 \leqslant k$. Thus, by the pigeonhole principle, there exist at least m copies of some pair (D, D') in $\mathcal{D}$. Therefore, $H = (D + D')^m \leqslant G$, so by Lemma 5.7, $S_n(G) \geqslant S_n(H) \geqslant 2^{n-1}$ for every $n \leqslant m$. ◻

Finally, we make the following observation.

Observation 5.9. *Let $k, \ell, n \in \mathbb{N}$, let G be an ordered graph on $[n]$, and suppose that G can be partitioned into k blocks of consecutive vertices, with each block ℓ-homogeneous. Then there exists an ordered graph H on $[n]$ with at most k homogeneous blocks, such that $G \triangle H$ is ℓ-empty.*

Proof. Let the ℓ-homogeneous blocks of G be $B_1, \ldots, B_k$, and let $x \in B_i$ and $y \in B_j$ with $i, j \in [k]$. Either all or none of the edges of length at least ℓ between B_i and B_j are in G. Let $xy \in E(H)$ if and only if all of these edges are in G. (So if there are no edges of length at least ℓ, then $xy \notin E(H)$.) Then $G \triangle H$ is ℓ-empty. ◻

6 The Structure of a Property with Speed $p(n)F_{n,\ell}$

We can now deduce the structure of every ordered graph $G \in \mathcal{P}$, if $\mathcal{P}$ is a hereditary property whose speed satisfies $|\mathcal{P}_n| < F_{n,\ell+1}$ for some $n, \ell \in \mathbb{N}$.

Theorem 6.1. *Let $n, \ell \in \mathbb{N}$, let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose that $|\mathcal{P}_n| < F_{n,\ell+1}$. Then there exist $k, k' \in \mathbb{N}$ such that every ordered graph $G \in \mathcal{P}$ is of the form $G = H \triangle J$, where H is an ordered graph with at most k homogeneous blocks, and $J \in \mathcal{J}(k',\ell)$.*

Proof. Let $n, \ell \in \mathbb{N}$, let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose that $|\mathcal{P}_n| < F_{n,\ell+1} \leqslant 2^{n-1}$. By Corollary 3.6, there exists an integer $K \in \mathbb{N}$ such that every ordered graph $G \in \mathcal{P}$ may be partitioned into at most K blocks of consecutive vertices with each block ℓ-homogeneous.

Let $G \in \mathcal{P}$, let B be an ℓ-homogeneous block of G, and let $F = G[B]$. Suppose that F is ℓ-empty. By Lemma 5.2, F has at most $n-1$ irreducible blocks of size at least $\ell+1$, and by Lemma 5.3, F has no irreducible block of size at least $4n(\ell+1)$, since $S_n(F) \leqslant |\mathcal{P}_n| < F_{n,\ell+1}$. Therefore, by deleting a set of edges which span at most $4n^2(\ell+1)$ vertices of F, we can obtain an ordered graph F' in which each irreducible block has size at most ℓ.

Now, by Lemma 5.6, at most $n-1$ of these irreducible blocks are not in $\mathcal{J}$, since $|\mathcal{P}_n| < 2^{n-1}$. So by deleting a set of edges which span at most $(n-1)\ell$ vertices from F', we can obtain an ordered graph F'' in which each irreducible block is in $\mathcal{J}_\ell$.

Finally, let $s \in \mathbb{N}$ be minimal such that $F'' \in \mathcal{J}(s,\ell)$. By Lemma 5.8 we have $s \leqslant 50n\ell^2$, since $|\mathcal{P}_n| < 2^{n-1}$.

We have shown that if B is an ℓ-homogeneous block of G which induces an ℓ-empty graph F, then by deleting a set of edges which span at most $4n^2(\ell+1)+(n-1)\ell < 9n^2\ell$ vertices of F, we can obtain an ordered graph in $\mathcal{J}(50n\ell^2,\ell)$. By symmetry, if B induces an ℓ-complete graph F^*, then by adding a set of non-edges which span at most $9n^2\ell$ vertices of F^*, we can obtain a graph whose complement is in $\mathcal{J}(50n\ell^2,\ell)$. For each ℓ-homogeneous block B_i of G, choose such a collection of edges, E_i. Let H' be the ordered graph on $[N] = V(G)$, with edge set $\bigcup_i E_i$, and note that H' is ℓ-empty.

Now, every ordered graph in $\mathcal{P}$ may be partitioned into K or fewer ℓ-homogeneous blocks, $B_1 < \ldots < B_K$. Thus, by Observation 5.9, there exists an ordered graph H'' on $[n]$ with at most K homogeneous blocks, such that $G \triangle H''$ is ℓ-empty.

Finally, we need to remove edges of $G\triangle H''$ between different ℓ-homogeneous blocks of G, so let H''' have vertex set $[N]$ and edge set $E(H') \cup \{uv \in E(G\triangle H'') : u \in B_i, v \in B_j, i \neq j\}$.

Let $H = H''\triangle H'''$. We claim that H has at most $18Kn^2\ell + 4K\ell + K$ homogeneous blocks, and that $G\triangle H \in \mathcal{J}(k',\ell)$ if $k' \geqslant 50Kn\ell^2$. The first statement follows because fewer than $9Kn^2\ell + 2K\ell$ vertices of H''' have non-zero degree, and H'' has at most K homogeneous blocks. To prove the second statement, note that $G\triangle H = (G\triangle H'')\triangle H'''$, so $G\triangle H = A_1 + \ldots + A_K$, where $A_i \in \mathcal{J}(50n\ell^2,\ell)$ for each $i \in [K]$. It is now easy to see that $G\triangle H \in \mathcal{J}(k',\ell)$ if $k' \geqslant 50Kn\ell^2$.

Thus, letting $k = 18Kn^2\ell + K + 4K\ell$ and $k' = 50Kn\ell^2$, we have $G = H\triangle J$ for some $J \in \mathcal{J}(k',\ell)$. ◻

7 Proof of Theorem 1.1

We shall use the following easy fact about Fibonacci numbers.

Observation 7.1. *Let $\ell, m, n \in \mathbb{N}$. Then $F_{m+n,\ell} \geqslant F_{m,\ell} \cdot F_{n,\ell}$.*

Proof. Let $\ell \in \mathbb{N}$. We use induction on $m+n$. We have $F_{2,\ell} = 2 > 1 = F_{1,\ell} \cdot F_{1,\ell}$, so the result holds for $m+n=2$. Now, let $m+n>2$, and assume the result is true for all smaller values of $m+n$. Then

$$\begin{aligned} F_{m+n,\ell} &= F_{m+n-1,\ell} + \ldots + F_{m+n-\ell,\ell} \\ &\geqslant F_{m,\ell}\left(F_{n-1,\ell} + \ldots + F_{n-\ell,\ell}\right) = F_{m,\ell} \cdot F_{n,\ell}. \end{aligned}$$

So the induction step holds, and the observation is proved. ◻

The following lemmas provide the final piece of the jigsaw.

Lemma 7.2. *Let $\ell, n \in \mathbb{N}$. Then $|\mathcal{J}(1,\ell)_n| = O(F_{n,\ell})$ as $n \to \infty$.*

Proof. Let $\ell, n \in \mathbb{N}$, and let $G \in \mathcal{J}(1,\ell)_n$. Then for some $t \in \mathbb{N}$ we have $G = B_1 + \ldots + B_t$, with $B_j \in \mathcal{J}_\ell$ for each $j \in [t]$, and $B_i \leqslant B_j$ or $B_j \leqslant B_i$ for each pair $i, j \in [t]$. This is the irreducible block decomposition $B(G)$ of G, and so is clearly unique. Note also that there is a unique 'largest' ordered graph B in $B(G)$, i.e., $B_j \leqslant B$ for each $j \in [t]$, and $B_i = B$ for some $i \in [t]$.

How many such ordered graphs G are there? Partition $\mathcal{J}(1,\ell)$ as follows: for each $B \in \mathcal{J}_\ell$, let $\mathcal{J}(B) = \{G \in \mathcal{J}(1,\ell) : B$ is the largest irreducible graph in $B(G)\}$. Each ordered graph $B \in \mathcal{J}_\ell$ has order at most ℓ, so there are only a bounded number of them. Hence we will be done if we can prove that $|\mathcal{J}(B)_n| = O(F_{n,\ell})$ for every $B \in \mathcal{J}_\ell$.

But this is now easy, since every ordered graph in $\mathcal{J}(B)$ is a subgraph of B^m for some sufficiently large m. Now simply observe that for each $B \in \mathcal{J}_\ell$, B has exactly one irreducible ordered subgraph of order n for each $n \leqslant |B|$, and it follows by a simple induction on n that $|\mathcal{J}(B)_n| = F_{n,|B|} \leqslant F_{n,\ell}$ for every $n \in \mathbb{N}$. □

Using Lemma 7.2, we can now give an upper bound on $|\mathcal{J}(k,\ell)_n|$ for all k, ℓ and $n \in \mathbb{N}$.

Lemma 7.3. *Let $k, \ell, n \in \mathbb{N}$. Then $|\mathcal{J}(k,\ell)_n| = O(n^{k-1} F_{n,\ell})$ as $n \to \infty$.*

Proof. Let $k, \ell, n \in \mathbb{N}$. If $G \in \mathcal{J}(k,\ell)_n$, then $G = A_1 + \ldots + A_k$, with each $A_i \in \mathcal{J}(1,\ell) \cup \emptyset$ (where $\emptyset$ here denotes the ordered graph with $|G| = 0$). Let $a(i) = |A_i|$ for each $i \in [k]$, so $a_i \in \mathbb{N} \cup \{0\}$.

It follows from this that an ordered graph $G \in \mathcal{J}(k,\ell)_n$ is determined by a sequence $(a(1), \ldots, a(k))$, with $a(i) \in \mathbb{N} \cup \{0\}$ for each $i \in [k]$, and $\sum_i a_i = n$; and a sequence of ordered graphs $(A_1, \ldots, A_k)$, with $A_i \in \mathcal{J}(1,\ell)_{a(i)}$ for each $i \in [k]$. There are $O(n^{k-1})$ such sequences of integers, and by Observation 7.1 and Lemma 7.2 there are

$$\prod_{i=1}^{k} O(F_{a(i),\ell}) = O(F_{a(1)+\ldots+a(k),\ell}) = O(F_{n,\ell})$$

such sequences of ordered graphs. The result follows immediately. □

Proof of Theorem 1.1. Let $\mathcal{P}$ be a hereditary property of ordered graphs, and suppose that $|\mathcal{P}_n| < 2^{n-1}$ for some $n \in \mathbb{N}$. In particular, let $m \in \mathbb{N}$ satisfy $|\mathcal{P}_m| < 2^{m-1}$. Since $F_{n,\ell} = 2^{n-1}$ for every $n \leqslant \ell$, it follows that $|\mathcal{P}_m| < F_{m,m}$. Let $\ell \in \mathbb{N}$ be the minimal integer such that $|\mathcal{P}_n| < F_{n,\ell+1}$ for some $n \in \mathbb{N}$; we have shown that $\ell \leqslant m - 1$, so such an integer ℓ exists. Note that by the definition of ℓ, $|\mathcal{P}_n| \geqslant F_{n,\ell}$ for every $n \in \mathbb{N}$.

Suppose first that $\ell = 1$, so $|\mathcal{P}_n| < F_n$ for some $n \in \mathbb{N}$. Theorem 4.6 and Lemma 4.2 then imply that either case (a) or case (b) of the theorem holds.

So let $\ell \geqslant 2$, and apply Theorem 6.1 to $\mathcal{P}$. By the theorem, there exist integers $k, k' \in \mathbb{N}$ such that every ordered graph $G \in \mathcal{P}$ may be written as $G = H \triangle J$, where H has at most $k+1$ homogeneous blocks, and $J \in \mathcal{J}(k'+1, \ell)$. By Lemma 4.1 there are $O(n^k)$ such ordered graphs H on n vertices, and by Lemma 7.3 there are $O(n^{k'} F_{n,\ell})$ such ordered graphs J on n vertices. Hence $|\mathcal{P}_n| = O(n^{k+k'} F_{n,\ell})$. ◻

8 Further Problems

Theorems 1.1 and 1.2 restrict the possible speeds of a hereditary property of ordered graphs if the speed is at most 2^{n-1}, or at least 2^{cn^2} for some $c > 0$. There are many obvious questions remaining in the large gap between these ranges, and in this section we shall discuss some of these.

For hereditary properties of both labelled graphs and permutations, there is a jump from exponential speed (speed c^n for some constant $c > 0$) to factorial speed (speed n^{cn} for some constant $c > 0$) (see [BBW00] and [MT04]). As we have seen, ordered graphs generalize both of these types of structure, so it is natural to ask whether a similar jump occurs for ordered graphs. In [BBM06$^+$b] it was proved that such a jump does occur for hereditary properties of ordered graphs in which every component is a clique, for monotone properties of ordered graphs (properties closed under taking arbitrary (i.e., not necessarily induced) ordered subgraphs), and for hereditary properties of ordered graphs not containing arbitrarily large complete, or complete bipartite ordered graphs. (The first two of these results have been proved independently by Klazar and Marcus [KM06$^+$].) It was also conjectured that the same jump holds for arbitrary hereditary properties of ordered graphs.

Even assuming a positive answer to the conjecture, one is still left with the problem of determining the possible exponential speeds. In particular, we have the following questions.

Conjecture 8.1. *If $\mathcal{P}$ is a hereditary property of ordered graphs, and $|\mathcal{P}_n| < c^n$ for some $c \in \mathbb{R}$ and every $n \in \mathbb{N}$, then $\lim_{n\to\infty} (|\mathcal{P}_n|)^{1/n}$ exists.*

Open Problem 8.2. Let $\mathcal{S} = \{c \in \mathbb{R}$: there is a hereditary property of ordered graphs $\mathcal{P}$ with $\lim_{n\to\infty} (|\mathcal{P}_n|)^{1/n} = c\}$. Determine the set $\mathcal{S}$.

Theorem 1.1 solves Problem 8.2 in the case $\liminf_{n\to\infty} (|\mathcal{P}_n|)^{1/n} < 2$. The following corollary is immediate from the theorem.

Corollary 8.3. *Let $\mathcal{S}$ be as defined in Problem 8.2, and let $A = \{x : x$ is the largest real root of the polynomial $x^{k+1} = x^k + x^{k-1} + \ldots + 1$ for some $k \in \mathbb{N}\}$. Then $\mathcal{S} \cap [0,2] = \{0,2\} \cup A$.*

Arratia [Arr99] proved Conjecture 8.1 for principal hereditary properties of permutations. The following easy result (which uses basically the same method) proves another special case of the conjecture.

Theorem 8.4. *Let $G_1, G_2, \dots$ be a sequence of ordered graphs, and suppose that either every G_i is irreducible, or every $\overline{G_i}$ is irreducible. Let $\mathcal{P} = \{G : G$ is an ordered graph, and $G_i \nleqslant G$ for every $i \in \mathbb{N}\}$. Then either $\lim_{n\to\infty} (|\mathcal{P}_n|)^{1/n}$ exists, or $\liminf_{n\to\infty} (|\mathcal{P}_n|)^{1/n} = \infty$.*

Proof. We claim that for every pair of integers m, n,

$$|\mathcal{P}_{n+m}| \geqslant |\mathcal{P}_n| \cdot |\mathcal{P}_m|.$$

Assume that every G_i is irreducible (the other case can be dealt with similarly). Let $F_1 \in \mathcal{P}^n$ and $F_2 \in \mathcal{P}^m$, and let $i \in \mathbb{N}$. By the definition of $\mathcal{P}$ we have $G_i \nleqslant F_1$ and $G_i \nleqslant F_2$, and so, since G_i is irreducible, $G_i \nleqslant F_1 + F_2$. This holds for every $i \in \mathbb{N}$, so $F_1 + F_2 \in \mathcal{P}^{n+m}$. This proves the claim.

Now, Fekete's Lemma [Fek23] states that if $a_1, a_2, \dots \in \mathbb{R}$ satisfy $a_m + a_n \geqslant a_{m+n}$ for all $m, n \geqslant 1$, then $\lim_{n\to\infty} \dfrac{a_n}{n}$ exists and is in $[-\infty, \infty)$. Applying this lemma to the sequence $-\log(|\mathcal{P}_n|)$ gives the result. □

There has been a large volume of work done on the possible exponential speeds of principal hereditary properties of permutations (see for example [Bóna04]). Until recently all such known speeds were of the form $k^{(1+o(1))n}$, with $k \in \mathbb{N}$, but a non-integer base was found by Bóna [Bóna05], who proved that the property of all permutations avoiding 12453 has speed $(9 + 4\sqrt{2})^{(1+o(1))n}$. However, there are no known hereditary properties with speed $c^{(1+o(1))n}$ and c transcendental. We have been unable to find even a hereditary property of ordered graphs with transcendental base, but the following simple construction shows that for properties of ordered graphs, irrational bases are much easier to come by than in the more restrictive principal permutation property setting.

Theorem 8.5. *Let $k \in \mathbb{N}$, and let $a(0) \leqslant \dots \leqslant a(k)$ with $a(i) \in \mathbb{N}$ for each i. Let c be the largest real root of the polynomial $x^{k+1} = \sum_{i=0}^{k} a(i)x^i$. Then there exists a hereditary property of ordered graphs $\mathcal{P}$ with $|\mathcal{P}_n| = c^{(1+o(1))n}$.*

Proof. Let $k \in \mathbb{N}$, and let $a(0) \leqslant \dots \leqslant a(k)$ with $a(i) \in \mathbb{N}$ for each $i \in [0, k]$. We shall define a particular infinite ordered graph G, and let $\mathcal{P}$ be the property of ordered graphs consisting of all (finite, order-preserving) subgraphs of G.

For each $\ell \in [k]$ let K_ℓ denote a copy of the complete ordered graph on ℓ vertices, and let $H = K_{k+1}^{a(0)} + K_k^{a(1)-a(0)} + \dots + K_1^{a(k)-a(k-1)}$. Let $BS(H) = (b(1), \dots, b(a(k)))$, for each $j \in [a(k)]$ let $d(j) = \sum_{i=1}^{j-1} b(i)$, and let $d = \sum_{i=1}^{a(k)} b(i) = |H|$. Let G be an infinite ordered graph with vertex set $\mathbb{N}$, satisfying the following conditions:

(*i*) $G - [a(k)] = H + H + \dots$, and
(*ii*) for each $i \in [a(k)]$ and each $a(k) < j$, $ij \in E(G)$ if and only if $j - a(k) \in [d(i) + 1, d] \pmod d$.

Let $\mathcal{P}$ be the collection of all (finite, order-preserving) induced subgraphs of G. It is easy to see that $\mathcal{P}$ is a hereditary property of ordered graphs. Let T_n be the sequence of integers defined by $T_n = 0$ if $n < 0$, $T_0 = 1$, and $T_{n+1} = \sum_{t=0}^{k} a(k-t)T_{n-t}$ for every $n \geqslant 0$. We claim that $|\mathcal{P}_n| = \Theta(T_n)$.

Let us first show that $|\mathcal{P}_{n+a(k)}| \geqslant T_n$ for every $n \in \mathbb{N}$. Indeed, let the first (leftmost) $a(k)$ vertices of G be denoted A, and let's consider only those subgraphs of G, on $n + a(k)$ vertices, which include all of A. Such an ordered graph consists of A, and then a sequence of cliques, each connected to a subset (in fact an initial segment) of A. If the clique has size t, then there are exactly $a(k+1-t)$ choices for this subset. It now follows easily by induction on n that there are exactly T_n such ordered subgraphs.

Now, note that any graph in $\mathcal{P}_n$ may be obtained by first taking an ordered subgraph J of G on $n+i$ vertices (with $0 \leqslant i \leqslant a(k)$) containing all of A, and then removing i vertices of A. Since T_n is increasing, there are at most T_n choices for J (for a given i), and hence $|\mathcal{P}_n| \leqslant 2^{a(k)}T_n$. This proves that $|\mathcal{P}_n| = \Theta(T_n)$, and the theorem follows. □

An *accumulation point from below* of a set $S \subset \mathbb{R}$ is a point $c \in \mathbb{R}$ such that for every $\varepsilon > 0$, $S \cap (c-\varepsilon, c) \neq \emptyset$. Let $A^1(S)$ denote the accumulation points from below of S, and for each $n \in \mathbb{N}$, let $A^{n+1}(S)$ denote the accumulation points from below of the set $A^n(S)$. We call the set $A^n(S)$ the *degree n accumulation points from below*. Using Theorem 8.5, we can obtain the following result about the accumulation points of $\mathcal{S}$.

Corollary 8.6. *For each $2 \leqslant n \in \mathbb{N}$ we have $n \in A^{n-1}(\mathcal{S})$. In other words, n is a degree $n-1$ accumulation point from below of $\mathcal{S}$.*

Proof. We claim that the result holds even if we consider only the family of properties described in Theorem 8.5, and prove this claim by induction on n. For $n = 2$, consider the sequences (1), $(1,1)$, $(1,1,1)$, and so on, and apply Theorem 8.5. This gives a sequence of constants $c_i \in \mathcal{S}$ with $c_i \to 2^-$ as $i \to \infty$ (note that in fact $c_i = \lim_{n\to\infty} (F_{n,i})^{1/n}$), so the claim is true for $n = 2$.

So let $3 \leqslant n \in \mathbb{N}$ and assume the claim holds for $n-1$. Consider the collection $\mathcal{A}$ of sequences which proved the result for $n-1$. Now add 1 to each entry of each sequence in $\mathcal{A}$, to obtain the collection $\mathcal{A}'$. The sequences in $\mathcal{A}'$ all still satisfy the conditions of Theorem 8.5, so we may apply the theorem to them. It follows that n is a degree $n-2$ accumulation point from below of $\mathcal{S}$.

Now, let $\omega = (a(1), \dots, a(k)) \in \mathcal{A}'$, and suppose applying Theorem 8.5 to ω shows that $c \in \mathcal{S}$. We sub-claim that $c \in A^1(\mathcal{S})$; since ω was arbitrary, this will suffice to prove the claim. Note that $a(k) \geqslant 2$, consider the sequences $(a(1), \dots, a(k), 1)$, $(a(1), \dots, a(k), 1, 1)$, and so on, and apply Theorem 8.5. The theorem gives a sequence of constants $c_i \in \mathcal{S}$ with $c_i \to c^-$ as $i \to \infty$, so $c \in A^1(\mathcal{S})$ as sub-claimed. Thus $n \in A^{n-2}(A^1(\mathcal{S})) = A^{n-1}(\mathcal{S})$, and the induction step is complete. The result follows immediately. □

Of course it is not necessary to use copies of the complete graph in the proof of Theorem 8.5 – one could use the irreducible graphs from any finite hereditary property of ordered graphs. However it does not appear that arbitrary sequences are possible (at least using this method). The following conjecture is motivated by Theorem 8.5 and by Lemma 5.5.

Conjecture 8.7. *The smallest $c \in \mathcal{S}$ with $c > 2$ is the largest real root of the polynomial $x^5 = x^4 + x^3 + x^2 + 2x + 1$, and is approximately* 2.03.

An example of a hereditary property with this speed is the following. Let $\mathcal{P}$ consist of all ordered graphs in which each irreducible block is either $J_2^{(n)}$ with $n \leqslant 5$, or Q_1. It is easy to show that $\mathcal{P}$ is hereditary and has the desired speed. The following conjecture is a much more general version of Conjecture 8.7. It says that there is a jump everywhere!

Conjecture 8.8. *For every $c \in \mathbb{R}$, there exists an $\varepsilon = \varepsilon(c)$ such that $\mathcal{S} \cap (c, c+\varepsilon) = \emptyset$. In particular, $\mathcal{S}$ has no accumulation points from above.*

All the members of $\mathcal{S}$ we have found are either integers or algebraic irrationals. Our final conjecture says that all members of $\mathcal{S}$ are of one of these two types.

Conjecture 8.9. *Every $c \in \mathcal{S}$ is either an integer or an irrational algebraic number, so $\mathcal{S} \subset \mathbb{Z} \cup (\mathbb{A} \setminus \mathbb{Q})$.*

Finally, a bipartite ordered graph is a bipartite graph together with a linear order on each part. Note that bipartite ordered graphs are equivalent to $\{0, 1\}$-matrices, and that a bipartite ordered graph may be mapped to an ordered graph by placing one part to the left of the other. It is easy to see that for bipartite ordered graphs, structures of Type 3 do not occur. We thus obtain the following corollaries to the proof of Theorem 1.1.

Corollary 8.10. *There exists a function $g : \mathbb{N} \to \mathbb{N}$ such that if $k \in \mathbb{N}$ and G is a bipartite ordered graph containing no k-structure of Type 1 or Type 2, then G can be partitioned into at most $g(k)$ homogeneous blocks.*

Corollary 8.11. *If $\mathcal{P}$ is a hereditary property of bipartite ordered graphs, then either*

(a) $|\mathcal{P}_n| = \sum_{i=0}^{k} a_i \binom{n}{i}$ *for some $k \in \mathbb{N}$, $a_0, \ldots, a_k \in \mathbb{Z}$ and all sufficiently large n, or*

(b) $|\mathcal{P}_n| \geqslant 2^{n-1}$ *for every $n \in \mathbb{N}$.*

Acknowledgement. We would like to thank the anonymous referees for their careful reading of the manuscript, and for numerous helpful comments.

References

[Ale93] V. E. Alekseev, On the entropy values of hereditary classes of graphs, *Discrete Math. Appl.* **3** (1993), 191–199.

[AS00] V. E. Alekseev, S. V. Sorochan, On the entropy of hereditary classes of oriented graphs, (Russian) International Conference DAOR 2000, *Diskretn. Anal. Issled. Oper. Ser. 1*, **7** (2000), 20–28.

[Arr99] R. Arratia, On the Stanley–Wilf conjecture for the number of permutations avoiding a given pattern, *Electron. J. Combin.*, **6** (1999), 1, 4pp.

[BB05] J. Balogh and B. Bollobás, Hereditary properties of words, *RAIRO Theoretical Informatics and Applications*, Special Issue (Imre Simon), **39** (2005), 49–66.

[BBM06$^+$a] J. Balogh, B. Bollobás and R. Morris, Hereditary properties of combinatorial structures: posets and oriented graphs, submitted to *J. Graph Theory.*

[BBM06$^+$b] J. Balogh, B. Bollobás and R. Morris, Hereditary properties of partitions, ordered graphs and ordered hypergraphs, to appear in a special edition of *Europ. J. Combin.*, eds. B. Sudakov and M. Krivelevich.

[BBM06$^+$c] J. Balogh, B. Bollobás and R. Morris, Hereditary properties of tournaments, in preparation.

[BBS04] J. Balogh, B. Bollobás and M. Simonovits, On the number of graphs without forbidden subgraph, *J. Combin. Theory Ser. B.*, **91** (2004), 1–24.

[BBW00] J. Balogh, B. Bollobás, and D. Weinreich, The speed of hereditary properties of graphs, *J. Combin. Theory Ser. B*, **79** (2000), 131–156.

[BBW06$^+$] J. Balogh, B. Bollobás and D. Weinreich, A jump to the Bell number for hereditary graph properties, *to appear in J. Combin. Theory Ser. B.*

[Bol98] B. Bollobás, Hereditary properties of graphs: asymptotic enumeration, global structure, and colouring, in *Proceedings of the International Congress of Mathematicians*, Vol. III (Berlin, 1998), *Doc. Math.* 1998, Extra Vol. III, 333–342 (electronic).

[BT95] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.*, **27** (1995) 417–424.

[BT97] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, “The mathematics of Paul Erdős, II” (R.L. Graham and J. Nešetřil, Editors), *Alg. and Combin.*, Vol. 14, Springer-Verlag, New York/Berlin (1997), 70–78.

[Bóna04] M. Bóna, Combinatorics of permutations. With a foreword by Richard Stanley. Discrete Mathematics and its Applications (Boca Raton). Chapman & Hall/CRC, Boca Raton, FL, 2004.

[Bóna05] M. Bóna, The limit of a Stanley-Wilf sequence is not always rational and layered patterns beat monotone patterns, *J. Combin. Theory*, **110** (2005), 223-235.

[BGP99] G. Brightwell, D. A. Grable and H. J. Prömel, Forbidden induced partial orders, *Discrete Math.*, **201** (1999), 53–90.

[Erd64] P. Erdős, On extremal problems of graphs and generalized graphs, *Israel J. Math.*, **2** (1964), 183–190.

[EFR86] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs and Combin.*, **2** (1986), 113–121.

[EKR76] P. Erdős, D. J. Kleitman and B. L. Rothschild, Asymptotic enumeration of K_n-free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie* (Rome, 1973), Vol. II, pp. 19–27. *Atti dei Convegni Lincei*, **17**, Accad. Naz. Lincei, Rome, 1976.

[Fek23] M. Fekete, Uber die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten, *Math. Z.*, **17** (1923) 228–249.

[HPS93] C. Hundack, H. J. Prömel and A. Steger, Extremal graph problems for graphs with a color-critical vertex. *Combin. Probab. Comput.*, **2** (1993), 465–477.

[KK03] T. Kaiser and M. Klazar, On growth rates of hereditary permutation classes, *Electr. J. Combinatorics*, **9** (2) (2003), 20pp.

[Kla00] M. Klazar, The Füredi–Hajnal conjecture implies the Stanley–Wilf conjecture, Formal Power Series and Algebraic Combinatorics (D. Krob, A. A. Mikhalev and A. V. Mikhalev, eds.), Springer, Berlin, (2000), 250–255.

[KM06+] M. Klazar and A. Marcus, Extensions of the linear bound in the Füredi–Hajnal conjecture, preprint.

[KPR87] Ph. G. Kolaitis, H. J. Prömel and B. L. Rothschild, K_{l+1}-free graphs: asymptotic structure and a 0-1 law, *Trans. Amer. Math. Soc.*, **303** (1987), 637–671.

[MT04] A. Marcus and G. Tardos, Excluded permutation matrices and the Stanley-Wilf conjecture, *J. Combin. Theory Ser. A*, **107** (2004), 153–160.

[PS92] H. J. Prömel and A. Steger, Excluding induced subgraphs III., A general asymptotic, *Random Structures Algorithms*, **3** (1992), 19–31.

[PS96] H. J. Prömel and A. Steger, On the asymptotic structure of sparse triangle free graphs, *J. Graph Theory*, **21** (1996), 137–151.

[PS96] H. J. Prömel and A. Steger, Counting H-free graphs, *Discrete Math.*, **154** (1996), 311–315.

[QZ04] A. Quas and L. Zamboni, Periodicity and local complexity, *Theoret. Comput. Sci.*, **319** (2004), 229–240.

[SZ94] E.R. Scheinerman and J. Zito, On the size of hereditary classes of graphs, *J. Combin. Theory Ser. B*, **61** (1994), 16–39.

A Proof and Generalizations of the Erdős–Ko–Rado Theorem Using the Method of Linearly Independent Polynomials

Zoltán Füredi*, Kyung-Won Hwang, and Paul M. Weichsel

[1] Department of Mathematics, University of Illinois, 1409 W. Green St., Urbana, IL 61801, USA, and Rényi Institute of Mathematics of the Hungarian Academy of Sciences, 1364 Budapest, P.O.Box 127, Hungary
z-furedi@math.uiuc.edu, furedi@renyi.hu

[2] Department of Mathematics, University of Illinois, 1409 W. Green St., Urbana, IL 61801, USA
khwang@math.uiuc.edu

[3] Department of Mathematics, University of Illinois, 1409 W. Green St., Urbana, IL 61801, USA
weichsel@math.uiuc.edu

Summary. Our aim is to exhibit a short algebraic proof for the Erdős–Ko–Rado theorem. First, we summarize the method of linearly independent polynomials showing that if $X_1, \dots, X_m \subset [n]$ are sets such that X_i does not satisfy any of the set of s intersection conditions R_i but X_i satisfies at least one condition in R_j for all $j > i$ then $m \le \binom{n}{s} + \binom{n}{s-1} + \cdots + \binom{n}{0}$. The EKR theorem is follows by carefully choosing the intersection properties and adding extra polynomials. We also prove generalizations for non-uniform families with various intersection conditions.

AMS Subject Classification. 05D05 and 05C35.
Keywords. Erdős–Ko–Rado, Intersecting family.

1 Proofs of the EKR Theorem

In 1961, Erdős, Ko, and Rado [EKR61] proved that if $\mathcal{F}$ is a k-uniform family of subsets of a set of n elements with $k \le \frac{1}{2}n$ and with every pair of members of $\mathcal{F}$ intersect, then $|\mathcal{F}| \le \binom{n-1}{k-1}$. They also showed that for $k < \frac{1}{2}n$, equality holds if $\mathcal{F}$ consists of all k-sets containing a given element of the underlying set.

* Research supported in part by the Hungarian National Science Foundation under grant OTKA NK 62321, and by the National Science Foundation under grant DMS 0140692.

In addition to their remarkable proof (induction on k and, for a given k, left-shifting and induction on n), there are many interesting new proofs. For example, in 1972, Katona [Kat72] used a simple and elegant argument, the permutation method. Daykin [Day74] obtained Erdős–Ko–Rado from the Kruskal–Katona theorem. Hajnal and Rothschild [HR73] proved it for $n > n_0(k)$ by an early version of the kernel (or delta-system) method, developed and used very successfully by Frankl [Fra77] (the first full description of the method was published in Deza, Erdős, and Frankl [DEF78]). The most remarkable technique was due to Lovász, in his ground-breaking paper [Lov79], which used a geometric representation to prove that the Shannon capacity of the Kneser graph $K(n,k)$ is at most $\binom{n-1}{k-1}$ for all $k \leq \frac{1}{2}n$, thus yielding another proof (and generalization). Wilson [Wil84] gave an ingenious proof, using Delsarte's linear programming bound. (Actually, he proved much more concerning t-intersecting families.) Finding different ways to prove EKR has been the subject of a set of papers recently by Ehud Friedgut from (1) Graph Homomorphisms [DF06$^+$] (a joint work with I. Dinur) and (2) Harmonic Analysis [Fri06$^+$].

One of the most powerful methods for counting the number of objects in a certain combinatorial structure is to correspond polynomials to the objects and to show that these polynomials are, in fact, linearly independent in some space. See, for example, Delsarte, Goethals and Seidel [DGS77] as well as Larman, Rogers and Seidel [LRS77] for deep early results. This method has been used to prove intersection theorems by Blokhuis [Blo90], and then by Alon, Babai and Suzuki [ABS91], and most recently by Ramanan [Ram97], Snevily [Sne03] and others. See the monograph of Babai and Frankl [BF92] for more details. Interestingly, none of the new algebraic proofs can be directly applied as a new proof for the original Erdős–Ko–Rado.

The aim of this paper is to exhibit a short algebraic proof for the Erdős–Ko–Rado theorem, and then to give a number of Frankl–Wilson–Ray–Chaudhuri type generalizations. Before that, in section 2, we summarize the essence of the polynomial method in a powerful lemma, in a form that best fits our purposes. In section 3 the new proof for EKR is given, in section 4 we summarize old generalizations for non-uniform hypergraphs, and finally in section 5 some new generalizations are presented.

2 The Polynomial Method for Intersection Theorems

Suppose that $\mathcal{A} = \{A_1, A_2, \ldots, A_m\}$ is a family of finite sets where each A_i is a subset of $[n] := \{1, 2, \ldots, n\}$. We say that the set X satisfies the intersection property (P, α) (where P is a set and α is a non-negative integer) if $|X \cap P| = \alpha$. Suppose that for each $A_i \in \mathcal{A}$ a set of (at most) s intersection properties are given

$$R_i := \{(P_{i1}, \alpha_{i1}), \ldots, (P_{is}, \alpha_{is})\}.$$

(An intersection condition can be repeated, even for the same i).

Lemma 2.1. *Suppose that for each $A_i \in \mathcal{A}$ one can find a set $X_i \subset [n]$ such that*

1. *X_i does not satisfy any of the conditions in R_i, and*
2. *X_i satisfies at least one condition in R_j for all $j > i$.*

Then

$$m \leq \binom{n}{s} + \binom{n}{s-1} + \cdots + \binom{n}{0}. \tag{1}$$

Proof. Define the (n-variable, real) polynomial $f_i(x)$ as follows.

$$f_i(x_1, x_2, \ldots, x_n) := \prod_{1 \leq u \leq s} \left(\left(\sum_{v \in P_{iu}} x_v \right) - \alpha_{iu} \right)$$

If we use the notation $\widehat{X}$ for the characteristic vector of $X \subset [n]$ (i.e., $\widehat{X}$ is a 0-1 vector from R^n with its t^{th} coordinate is 1 if and only if $t \in X$), then the scalar product $\widehat{X}\widehat{Y}$ is equal to $|X \cap Y|$ for all $X, Y \subset [n]$. With this notation one can rewrite f_i as

$$f_i\left(\widehat{X}\right) = \prod_u \left(\widehat{X}\widehat{P_{iu}} - \alpha_{iu}\right) = \prod_u \left(|X \cap P_{iu}| - \alpha_{iu}\right).$$

In any case, it is obvious that our conditions imply that

$$f_j\left(\widehat{X_i}\right) \begin{cases} = 0 \text{ if } i < j \\ \neq 0 \text{ if } i = j. \end{cases} \tag{2}$$

Now define the (integer coefficient, real, n-variable) multilinear polynomial g_i by eliminating all the parentheses from f_i and repeatedly replacing a higher order factor x_v^2 by x_v (for all $1 \leq v \leq n$). Note that for a 0-1 vector x one has $f_i(x) = g_i(x)$, so (2) implies that

$$g_i\left(\widehat{X_i}\right) \neq 0 \quad \text{but} \quad g_j\left(\widehat{X_i}\right) = 0 \text{ for } i < j. \tag{3}$$

The multilinear, n variable, real polynomials of degree at most s form a vectorspace V (over R), of dimension

$$\dim V = \binom{n}{s} + \binom{n}{s-1} + \cdots + \binom{n}{0}. \tag{4}$$

We claim that the polynomials $g_1, g_2, \ldots, g_m$ are linearly independent in this space. Then (4) gives the desired upper bound on m. Indeed, suppose on the contrary, that there exists a linear dependency, i.e.,

$$c_1 g_1(x) + c_2 g_2(x) + \cdots + c_m g_m(x) = 0 \tag{5}$$

holds for every $x \in R^n$, where the c_i's are reals, not all 0. Suppose that i is the smallest integer with $c_i \neq 0$, and substitute $\widehat{X_i}$ into (5). Since $c_j = 0$ for $j < i$ and $g_j\left(\widehat{X_i}\right) = 0$ for $j > i$ the equation becomes $c_i g_i\left(\widehat{X_i}\right) = 0$. However, this contradicts the fact that both c_i and $g_i\left(\widehat{X_i}\right)$ are non-zero. □

3 The Polynomial Proof of EKR Theorem

Let $\mathcal{F}$ be an intersecting family of k-sets of $[n]$, $n \geq 2k$, $|\mathcal{F}| = m$. In the above Lemma 2.1 we will define $s = k - 1$. The dimension of the vector space of multilinear polynomials V is not a perfect binomial coefficient, so at first Lemma 2.1 does not seem to be applicable to prove $m \leq \binom{n-1}{k-1}$. Instead of narrowing the vectorspace, which does not look viable, one can add more polynomials to the g_i's defined by the members of $\mathcal{F}$, and show that the larger system is still linearly independent. This method appeared first in a paper of Blokhuis [Blo84]. In fact, we will join to $\mathcal{F}$ another

$$\binom{n}{k-1}+\binom{n}{k-2}+\cdots+\binom{n}{1}+\binom{n}{0}-\binom{n-1}{k-1} = 2\times\left(\sum_{k-2\geq u\geq 0}\binom{n-1}{u}\right) \tag{6}$$

sets, together with appropriate intersection conditions.

Select a member p of $[n]$ arbitrarily. Define the family $\mathcal{A}$ as the union of the following four hypergraphs, $\mathcal{F}_0$, $\mathcal{H}$, $\mathcal{F}_1$ and $\mathcal{G}$, where

- $\mathcal{F}_0 := \{F \in \mathcal{F} : p \notin F\}$,
- $\mathcal{H} := \{H \subset [n] : p \notin H,\ 0 \leq |H| \leq k-2\}$,
- $\mathcal{F}_1 := \{F \in \mathcal{F} : p \in F\}$, and
- $\mathcal{G} := \{G \subset [n] : p \in G,\ 1 \leq |G| \leq k-1\}$.

Order these sets linearly in the above order. First, we put the members of $\mathcal{F}_0$ (in arbitrary order), then the members of $\mathcal{H}$ increasing by size, (i.e., $H \in \mathcal{H}$ precedes $H' \in \mathcal{H}$ if $|H| \leq |H'|$) then $\mathcal{F}_1$ (again in arbitrary order) and, finally, the members of $\mathcal{G}$ again in increasing order. Next, for each set $A_i \in \mathcal{A}$ we associate another set $X_i \subset [n]$, and at most $k-1$ intersection conditions (P_{iu}, α_{iu}).

- For $F \in \mathcal{F}_0$ we let $X := [n] \setminus \{p\} \setminus F$ with intersection conditions (F, α), $1 \leq \alpha \leq k-1$.
- For $H \in \mathcal{H}$ we let $X := H$ with intersection conditions $(\{h\}, 0)$, (for each $h \in H$) and $([n], n-k-1)$.
- For $F \in \mathcal{F}_1$ we let $X := F \setminus \{p\}$ with intersection conditions $(F \setminus \{p\}, \alpha)$, $0 \leq \alpha \leq k-2$.
- For $G \in \mathcal{G}$ we let $X := G$ with intersection conditions $(\{g\}, 0)$ for each $g \in G$.

It is straightforward to check that the $\{A_i, X_i, (P_{iu}, \alpha_{iu})\}$ system defined above indeed satisfies the constrains of Lemma 2.1 with $s = k-1$. This gives $|\mathcal{F}|+|\mathcal{G}|+|\mathcal{H}| \le \binom{n}{k-1}+\binom{n}{k-2}+\cdots+\binom{n}{0}$. since $|\mathcal{G}| = |\mathcal{H}| = \sum_{k-2\ge u\ge 0}\binom{n-1}{u}$ (6) gives the desired upper bound for $|\mathcal{F}|$. ☐

Summary

To make the construction more explicit and transparent we repeat the the definitions of the associated functions and their characteristic properties.

For $F \in \mathcal{F}_0$ we have $f_F(x) = \prod_{1\le u\le k-1}\left(\left(\sum_{e\in F} x_e\right)-u\right)$ and $g_F\left(\widehat{X}\right)=0$ if and only if $|X\cap F| \in \{1,\dots,k-1\}$,

for $H \in \mathcal{H}$ we have $f_H(x) = \left(\left(\sum_{1\le e\le n} x_e\right)-(n-k-1)\right)\prod_{e\in H} x_e$ and $g_H\left(\widehat{X}\right)=0$ if and only if $|X| = n-k-1$ or $H \not\subseteq X$,

for $F \in \mathcal{F}_1$ we have $f_F(x) = \prod_{0\le u\le k-2}\left(\left(\sum_{e\in F, e\ne p} x_e\right)-u\right)$ and $g_F\left(\widehat{X}\right)=0$ if and only if $|X\cap(F\setminus\{p\})| \in \{0,1,\dots,k-2\}$,

for $G \in \mathcal{G}$ we have $f_G(x) = \prod_{e\in G} x_e$ and $g_G\left(\widehat{X}\right)=0$ if and only if $G \not\subseteq X$.

Let us point out the part where the condition $n \ge 2k$ was used. We needed $g_H\left(\widehat{H}\right) \ne 0$ and this is only true if $|H| \ne n-k-1$. Furthermore, as in most cases when one uses linear algebra, it does not seem immediate from the above proof that for $n > 2k$ equality can hold only for $\bigcap \mathcal{F} \ne \emptyset$.

4 Generalizations for Non-uniform Families

Here we briefly discuss some old generalizations. In this section $\mathcal{F} = \{F_1, F_2, \dots F_m\}$ is a family of subsets of $[n]$, $K = \{k_1,\dots,k_t\}$ and $L = \{l_1,\dots,l_s\}$ are sets of non-negative integers. We call $\mathcal{F}$ an (n,K,L)-family if $|F| \in K$ for every $F \in \mathcal{F}$ and $|F_i \cap F_j| \in L$ for any distinct $F_i, F_j \in \mathcal{F}$.

A celebrated result of Frankl and Wilson [FW81] claims that

$$|\mathcal{F}| \le \binom{n}{s}+\binom{n}{s-1}+\cdots+\binom{n}{0}, \tag{7}$$

independent of K. The determination of $\max|\mathcal{F}|$ is very much related to a basic coding problem when a binary code is given with given weights and distances. There are many improvements of (7).

For $L = \{1,2,\dots,s\}$ Frankl and Füredi [FF81] conjectured that

$$m \le \binom{n-1}{s}+\binom{n-1}{s-1}+\cdots+\binom{n-1}{0}. \tag{8}$$

This can be achieved by the family $\{F : 1 \in F \subset [n], |F| \le s+1\}$. They proved (8) for $n > 100s^2/\log(s+1)$, Pyber showed it for $n \ge 6(s+1)$ and finally Ramanan [Ram97] proved it for all n. Recently, Snevily [Sne03] showed that (8) holds for any L with $\min L \ge 1$.

Alon, Babai, and Suzuki [ABS91] proved that

$$m \le \binom{n}{s} + \binom{n}{s-1} + \binom{n}{s-t+1} \tag{9}$$

holds for the case $\min K > s - t$.

Snevily [Sne95] proved that in the case $\min K > \max L$

$$|\mathcal{F}| \le \binom{n-1}{s} + \binom{n-1}{s-1} + \cdots + \binom{n-1}{s-2t+1}.$$

Frankl [Fra85] showed that if $\mathcal{F}$ is any family with $|F_j \setminus F_i| \in L$ for all $1 \le i < j \le m$ then

$$|\mathcal{F}| \le \sum_{0 \le i \le s} \binom{n}{i}. \tag{10}$$

5 New Generalizations

Here we prove other generalizations of the Erdős–Ko–Rado Theorem allowing different sizes of the subsets and introducing a new parameter r.

Again $\mathcal{F} = \{F_1, F_2, \ldots, F_m\}$ is a family of subsets of $[n]$, $K = \{k_1, \ldots, k_t\}$ and $L = \{l_1, \ldots, l_s\}$ are sets of non-negative integers. We suppose that $|F| \in K$ for all $F \in \mathcal{F}$. Define

$$r := \min_{q \in [n]} \Big| \{|F| : q \in F \in \mathcal{F}\} \Big|,$$

the number of different sizes through q. Obviously, $r \le t$. Choose a vertex $p \in [n]$ so that $\Big|\{|F| : p \in F \in \mathcal{F}\}\Big| = r$. Assume, without loss of generality, that $p \notin F_i$ for $1 \le i \le a$ and $p \in F_i$ for $a < i \le |\mathcal{F}|$. Thus $\mathcal{F}$ has been split into two parts, $\mathcal{F}_0 = \{F : p \notin F \in \mathcal{F}\}$, and $\mathcal{F}_1 = \{F : p \in F \in \mathcal{F}\}$.

Theorem 5.1. *Suppose that the F_i's satisfy the following intersection properties:*

(i) $|F_j \setminus F_i| \in L$ *for all* $1 \le i < j \le a$,
(ii) $1 \le |F_j \setminus F_i| \le s$ *for* $1 \le i \le a < j \le m$,
(iii) $|F_i \cap F_j| \in \{1, \ldots, s\}$ *for* $a < i < j \le m$.

Suppose further that $\min L > 0$, $|F_i| > s$ *for* $a < i \le m$, *and* $n - k_i - 1 > s - r$ *for* $1 \le i \le r$. *Then*

$$|\mathcal{F}| \le \sum_{i=s-r+1}^{s} \binom{n-1}{i}.$$

In the proof of this theorem we define an n-variable multilinear polynomial f_F of degree at most s for each $F \in \mathcal{F}$. These polynomials are linearly independent (due to the intersection conditions) so we get an upper bound on $|\mathcal{F}|$. To decrease this bound we add more polynomials to this collection which form a space perpendicular to the space spanned by $\{f_F : F \in \mathcal{F}\}$, and then calculate the dimension of the spaces they span. The additional polynomials are defined using the families $\mathcal{H}$ and $\mathcal{G}$, the subsets $[n]$ of size at most $s-1$.

Proof. We proceed exactly like in Section 3. We divide the proof into two cases: $r \le s$ and $s < r$. First, we consider $r \le s$. Define the families

$\mathcal{H} = \{H \subseteq [n] : p \notin H, 0 \le |H| \le s-r\}$ and

$\mathcal{G} = \{G \subseteq [n] : p \in G, 1 \le |G| \le s\}$.

Order both of them linearly in increasing order, (i.e., H precedes H' if $|H| \le |H'|$). Put the four families $\mathcal{F}_0$, $\mathcal{H}$, $\mathcal{F}_1$ and $\mathcal{G}$ in this order (keeping their inner order). For each member A of these four families we associate an n variable polynomial $f_A(x_1, \ldots, x_n)$ and a special set X such that for the characteristic vector of X we have $f_A\left(\widehat{X}\right) \neq 0$ but $f_B\left(\widehat{X}\right) = 0$ for all B following A in the linear order. Reduce f_A, as it was done in Lemma 1, to a multilinear polynomial g_A by replacing every higher order term x_u^2 by x_u. We show that these polynomials are linearly independent, implying the upper bound for $|\mathcal{F}|$.

For each $F_i \in \mathcal{F}_0$, consider the polynomial

$$f_i(x) = \prod_{j=1}^{s} (v_i \cdot x - l_j),$$

where v_i is the characteristic vector of F_i. The special set is $[n] \setminus F_i \setminus \{p\}$.

For each $H \in \mathcal{H}$, we define the polynomial

$$f_H(x) = \prod_{b=1}^{r} \Big(\sum_{l=1}^{n} x_l - (n - k_b - 1) \Big) \prod_{j \in H} x_j.$$

The corresponding special set is H itself. Since $n - k_i - 1 > s - r$, $f_H\left(\widehat{H}\right) \neq 0$, and $f_{H'}\left(\widehat{H}\right) = 0$ for any $|H| \le |H'|$. Thus $\{f_H(x) : H \in \mathcal{H}\}$ is a linearly independent family.

For each $F_i \in \mathcal{F}_1$ let

$$f_i(x) = \prod_{j=0}^{s-1} (v_i^* \cdot x - j),$$

where v_i^* is the characteristic vector of $F_i \setminus \{p\}$. The special set is $F_i \setminus \{p\}$.

For each $G \in \mathcal{G}$ let

$$f_G(x) = \prod_{j \in G} x_j,$$

with special vector $\widehat{G}$. Note that $f_G\left(\widehat{G}\right) \neq 0$ and $f_{G'}\left(\widehat{G}\right) = 0$ for any $|G| \leq |G'|$, and thus $\{f_G(x) : G \in \mathcal{G}\}$ is a linearly independent family of polynomials.

For the rest of the proof of this case (i.e., the case $r \leq s$) we continue as in the proof of the EKR theorem.

For the $r > s$ case, we only need the family $\mathcal{G}$ as above, and then we show that $\{f_i(x) : F_i \in \mathcal{F}\} \cup \{f_G(x) : G \in \mathcal{G}\}$ is linearly independent. We get the bound $|\mathcal{F}| \leq \sum_{i=0}^{s} \binom{n-1}{i}$. (Note that the summation index i starts at 0 because $s - r + 1 \leq 0$). □

We give another generalization. Note that this result is almost identical to the previous one, but each is independent of the other (i.e., neither of them implies the other).

Theorem 5.2. *Suppose that with given n, K, L and $p \in [n]$ the family $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1$ satisfies the following intersection conditions*

(i) $|F_j \setminus F_i| \in L$ *for all* $1 \leq i < j \leq a$,
(ii) $1 \leq |F_j \setminus F_i| \leq s - r + 1$ *for* $1 \leq i \leq a < j \leq m$,
(iii) $|F_i \cap F_j| \in \{1, \ldots, s - r + 1\}$ *for* $a < i < j \leq m$.

Suppose further that $\min L > 0$, $|F_i| > s-r+1$ *for* $a < i \leq m$, *and* $n - k_i - 1 > s - r$ *for* $1 \leq i \leq r$. *Then*

$$|\mathcal{F}| \leq \sum_{i=s-r+1}^{s} \binom{n-1}{i}.$$

Proof. The proof closely follows the proof of Theorem 2 above. We divide the proof into the same two cases: $r \leq s$ and $s < r$. For the first case, we construct the same two families ($\mathcal{G}$ and $\mathcal{H}$) and the same associated polynomials, and for the second case we only use the family $\mathcal{G}$. The only difference is that the polynomial f_i for the set $F_i \in \mathcal{F}_1$ is defined as follows.

$$f_i(x) = \prod_{j=0}^{s-r} \left(v_i^* \cdot x - j\right).$$ □

Finally, the above method also gives the following bound for the general Erdős–Ko–Rado problem. For exact upper bound see Ahlswede and Khachatrian [AK97].

Theorem 5.3. *Let $\mathcal{F} = \{F_1, ..., F_m\}$ denote a family of k-subsets of $[n]$. Suppose that $|F_i \cap F_j| \in \{d, d+1, \cdots, k-1\}$ for all $i < j$. If $n \geq 2k - d + 1$, then $|\mathcal{F}| \leq \binom{n-1}{k-d}$.*

Acknowledgement. The authors are very grateful to the referee for many helpful advice.

References

[AK97] R. Ahlswede and L. H. Khachatrian, The complete intersection theorem for systems of finite sets. *European J. Combin.* **18** (1997), 125–136.

[ABS91] N. Alon, L. Babai, and H. Suzuki, Multilinear polynomials and Frankl–Ray–Chaudhuri–Wilson type intersection theorems. *J. Combin. Theory, Ser. A*, **58** (1991), 165–180.

[BF92] L. Babai and P. Frankl, *Linear Algebra Methods in Combinatorics.* (manuscript) Chicago University 1992.

[Blo84] A. Blokhuis, A new upper bound for the cardinality of 2-distance sets in Euclidean space. *Convexity and graph theory* (Jerusalem, 1981), 65–66, *North-Holland Math. Stud.* **87**, North-Holland, Amsterdam, 1984.

[Blo90] A. Blokhuis, Solution of an extremal problem for sets using resultants of polynomials. *Combinatorica* **10** (1990), 393–396.

[Day74] D. Daykin, Erdős–Ko–Rado from Kruskal–Katona. *J. Combinatorial Theory Ser. A* **17** (1974), 254–255.

[DGS77] P. Delsarte, J. M. Goethals, J. J. Seidel, Spherical codes and designs. *Geometriae Dedicata* **6** (1977), 363–388.

[DEF78] M. Deza, P. Erdős, and P. Frankl, Intersection properties of systems of finite sets. *Combinatorics* (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. I, pp. 251–256, *Colloq. Math. Soc. János Bolyai*, **18** North-Holland, Amsterdam-New York, 1978.

[DF83] M. Deza and P. Frankl, Erdős–Ko–Rado theorem—22 years later. *SIAM J. Algebraic Discrete Methods* **4** (1983), 419–431.

[DF06$^+$] I. Dinur and E. Friedgut, Proof of an intersection theorem via graph homomorphisms, to appear

[EKR61] P. Erdős, Chao Ko, and R. Rado, Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford* **12** (1961), 313–320.

[Fra77] P. Frankl, On families of finite sets no two of which intersect in a singleton. *Bull. Austral. Math. Soc.* **17** (1977), 125–134.

[Fra85] P. Frankl, Bounding the size of a family knowing the cardinality of differences. *Studia Sci. Math. Hungar.* **20** (1985), 33–36.

[FF81] P. Frankl and Z. Füredi, Families of finite sets with missing intersections. *Colloq. Math. Soc. János Bolyai* **37** (1981), 305–320.

[FW81] P. Frankl and R. M. Wilson, Intersection theorems with geometric consequences. *Combinatorica* **1** (1981), 357–368.

[Fri06$^+$] E. Friedgut, On the measure of intersecting families, uniqueness and robustness, to appear.

[HR73] A. Hajnal and B. Rothschild, A generalization of the Erdős–Ko–Rado theorem on finite set systems, *J. Combinatorial Theory Ser. A* **15** (1973), 359–362.

[Kat72] G. O. H. Katona, A simple proof of the Erdős–Ko–Rado theorem, *J. Combin. Theory, Ser. B* **13** (1972), 183–184.

[LRS77] D. G. Larman, C. A. Rogers, and J. J. Seidel, On two-distance sets in Euclidean space, *Bull. London Math. Soc.* **9** (1977), 261–267.

[Lov79] L. Lovász, On the Shannon capacity of a graph. *IEEE Trans. Inform. Theory* **25** (1979), 1–7.

[Pyb84] L. Pyber, An extension of a Frankl–Füredi theorem, *Discrete Math.* **52** (1984), 253–268.

[Ram97] G. Ramanan, Proof of a conjecture of Frankl and Füredi, *J. Combin. Theory, Ser. A* **79** (1997), 53-67.

[Sne95] H. Snevily, A generalization of the Ray–Chaudhuri–Wilson theorem, *J. Combin. Designs* **3** (1995), 349–352.

[Sne03] H. Snevily, A sharp bound for the number of sets that pairwise intersect at k positive values, *Combinatorica* **23** (2003), 527–532.

[Wil84] R. Wilson, The exact bound in the Erdős–Ko–Rado theorem, *Combinatorica* **4** (1984), 247–257.

Unions of Perfect Matchings in Cubic Graphs

Tomáš Kaiser[1*], Daniel Král'[2**], and Serguei Norine[3***]

[1] Department of Mathematics and Institute for Theoretical Computer Science (ITI), University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic
kaisert@kma.zcu.cz

[2] Institute for Theoretical Computer Science (ITI), Faculty of Mathematics and Physics, Charles University, Malostranské náměstí 25, 118 00 Prague, Czech Republic
kral@kam.mff.cuni.cz

[3] School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
snorine@math.gatech.edu

Summary. We show that any cubic bridgeless graph with m edges contains two perfect matchings that cover at least $3m/5$ edges, and three perfect matchings that cover at least $27m/35$ edges.

AMS Subject Classification. 05C70.
Keywords. Cubic graphs, perfect matchings, Berge-Fulkerson's conjecture.

1 Introduction

A well-known conjecture of Berge and Fulkerson states that every bridgeless cubic graph contains a family of six perfect matchings covering each edge exactly twice:

Conjecture 1.1. *Every cubic bridgeless graph G contains six perfect matching $M_1, \ldots, M_6$ such that each edge of G is contained in precisely two of the matchings.*

* Supported by the project 1M0545 and the Research Plan MSM 4977751301 of the Czech Ministry of Education.

** The author was a postdoctoral Institute for Mathematics, Technical University Berlin, Strasse des 17. Juni 136, D-10623 Berlin, Germany within the framework of the European training network COMBSTRU from October 2004 to July 2005. At the present, the author is a Fulbright scholar at School of Mathematics, Georgia Institute of Technology, 686 Cherry St, Atlanta, GA 30332-0160.

*** This author was partially supported by NSF Grant No. 0200595.

Conjecture 1.1 is attributed to Berge in [Sey79], but it was first published in [Ful71]. Closely related is the Cycle Double Conjecture (in particular, Conjecture 1.1 implies that every graph contains a family of cycles containing each edge four times). Note also that Conjecture 1.1 trivially holds for cubic graphs G that are 3-edge-colorable.

The following weaker version of Conjecture 1.1 due to Berge is also open:

Conjecture 1.2. *Every cubic bridgeless graph G contains five perfect matchings $M_1, \ldots, M_5$ such that each edge of G is contained in at least one of the matchings.*

We remark that even if the number 5 in Conjecture 1.2 is replaced by any larger constant (independent of G), the statement is not known to be true. In this paper, we investigate the maximum possible size of the union of a given number of perfect matchings in a cubic bridgeless graph. More precisely, we study, for $k \in \{2, 3\}$, the numbers

$$m_k = \inf_G \max_{M_1,\ldots,M_k} \frac{|\bigcup_i M_i|}{|E(G)|},$$

where the infimum is taken over all bridgeless cubic graphs G, and $M_1, \ldots, M_k$ range over all perfect matchings of G. Note that Conjecture 1.2 asserts that $m_5 = 1$.

We determine the precise value of m_2 and provide a non-trivial lower bound on m_3. Let us begin by considering the upper bounds. The Petersen graph P_{10} has 15 edges and 6 distinct perfect matchings. It can be checked that up to automorphism, the mutual position of any two (three) distinct perfect matchings in P_{10} is as shown in Figure 1a (Figure 1b, respectively). It follows that any two distinct perfect matchings in P_{10} have precisely one edge in common and the intersection of any three distinct perfect matchings is empty. Simple counting then shows that $m_2 \leq 3/5$ and $m_3 \leq 4/5$.

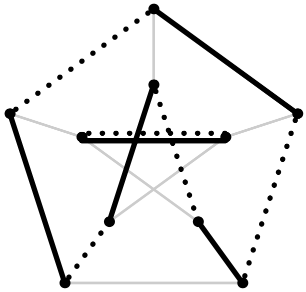
Our contribution can be summarized as follows:

Theorem 1.3. *The value of m_2 is $3/5$, and $0.771 \approx 27/35 \leq m_3 \leq 4/5$.*

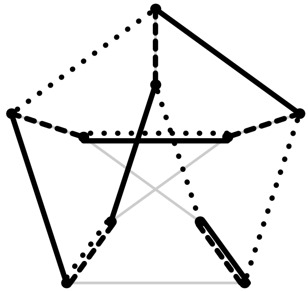
Our main tool is the Perfect Matching Polytope Theorem of Edmonds [Edm65] which we review in Section 2. Throughout the text, we use standard terminology and notation of graph theory as it can be found, e.g., in [Die00]. Supplementary information on Conjecture 1.1 can be found in [Zha97], and a more detailed introduction to the theory of matching polytopes can be found in the recent monograph by Schrijver [Sch03].

2 The Perfect Matching Polytope

Let $G = (V, E)$ be a graph which may contain multiple edges. A *cut* in G is any set $C \subseteq E$ such that $G \setminus C$ has more components than G does, and C is



(a)



(b)

Fig. 1. (a) Two perfect matchings in the Petersen graph P_{10} (solid and dotted edges) (b) Three perfect matchings in P_{10} (solid, dashed and dotted edges)

inclusion-wise minimal with this property. A k-cut (where k is an integer) is a cut comprised of k edges. For a subset $X \subseteq V$ of the vertices of G, we set ∂X to be the set of edges with precisely one end in X. Note that $\partial\emptyset = \partial V = \emptyset$. Finally, a cut C is said to be *odd* if there exists a subset $X \subseteq V$ of odd cardinality such that $C = \partial X$.

Let w be a vector in $\mathbb{R}^E$. The entry of w corresponding to $e \in E$ is denoted by $w(e)$, and for $A \subseteq E$, we define the *weight* $w(A)$ of A as $\sum_{e\in A} w(e)$. The vector w is said to be a *fractional perfect matching* of G if it satisfies the following:

(i) $0 \leq w(e) \leq 1$ for each $e \in E$,
(ii) $w(\partial\{v\}) = 1$ for each vertex $v \in V$, and
(iii) $w(\partial X) \geq 1$ for each $X \subseteq V$ of odd cardinality.

$P(G)$ denotes the set of all fractional perfect matchings of G. Let us remark that the third condition is dismissed in some literature (e.g., [SU97]).

If M is a perfect matching, then the characteristic vector $\chi^M \in \mathbb{R}^E$ of M is contained in $P(G)$. Furthermore, if $w_1, \ldots, w_n \in P(G)$, then any convex combination $\sum_{i=1}^n \alpha_i w_i$ of $w_1, \ldots, w_n$ (where $\alpha_1, \ldots, \alpha_n > 0$ are positive reals summing up to 1) also belongs to $P(G)$. It follows that $P(G)$ contains the convex hull of all the vectors χ^M where M is a perfect matching of G. The Perfect Matching Polytope Theorem asserts that the converse inclusion also holds:

Theorem 2.1 (Edmonds [Edm65]). *For any graph G, the set $P(G)$ coincides with the convex hull of the characteristic vectors of all perfect matchings of G.*

Rather naturally, $P(G)$ is called the *perfect matching polytope* of a graph G. Another fact that will be useful in our considerations is the following:

Lemma 2.2. *If w is a fractional perfect matching in a graph $G = (V, E)$ and $c \in \mathbb{R}^E$, then G has a perfect matching M such that*

$$c \cdot \chi^M \geq c \cdot w,$$

where $\cdot$ denotes the scalar product. Moreover, there exists such a perfect matching M that contains exactly one edge of each odd cut C with $w(C) = 1$.

Proof. Since w is a fractional perfect matching, it can be expressed by Theorem 2.1 as a convex combination of characteristic vectors of perfect matchings $M_1, \ldots, M_k$ of G. Since the cut C is odd, $\chi^{M_i}(C) \geq 1$ for every $i = 1, \ldots, k$. On the other hand, since w is a convex combination of $\chi^{M_1}, \ldots, \chi^{M_k}$ and $w(C) = 1$, it must hold that $\chi^{M_i}(C) = 1$ for every $i = 1, \ldots, k$. Since $c \cdot w$ is a convex combination of $c \cdot \chi^{M_i}$, it holds that $c \cdot \chi^{M_i} \geq c \cdot w$ for at least one of the perfect matchings. □

3 Proof of Theorem 1.3

In this section, we prove Theorem 1.3. By our discussion in Section 1, it suffices to show that $m_2 \geq 3/5$ and $m_3 \geq 27/35$:

Proof of Theorem 1.3. Fix a cubic bridgeless graph G. Define $w_1 \in \mathbb{R}^E$ to have the value $1/3$ on all edges $e \in E$. It is easy to verify that w_1 is a fractional perfect matching of G: Clearly, the conditions (i) and (ii) hold for every vertex v of G. Since G is bridgeless, every odd cut of G is comprised of at least 3 edges and thus condition (iii) also holds. Note that $w_1(C) = 1$ for each 3-cut C of G. Observe further that since G is cubic, all 3-cuts in G are odd (and more generally, a cut C in G is odd if and only if $|C|$ is odd). Hence, by Lemma 2.2, there is a perfect matching M_1 intersecting each 3-cut in a single edge. We remark that the existence of M_1 can be alternatively shown by induction on the order of G.

We now use M_1 to define the following vector $w_2 \in \mathbb{R}^E$:

$$w_2(e) = \begin{cases} 1/5 & \text{if } e \in M_1, \\ 2/5 & \text{otherwise.} \end{cases}$$

We verify that w_2 is a fractional perfect matching of G: the conditions (i) and (ii) clearly hold. Let C be an odd cut of G. The size of C is odd and it is at least 3. If C is a 3-cut, then $w(e) = 1/5$ for exactly one of the edges e contained in C and $w(e) = 2/5$ for the remaining two edges. If the size of C is 5 or more, then $w(C) \geq 5 \cdot 1/5 = 1$. Hence, condition (iii) also holds.

For each $e \in E$, set $c_2(e) = 1 - \chi^{M_1}(e)$. By Lemma 2.2, there exists a perfect matching M_2 such that

$$c_2 \cdot \chi^{M_2} \geq c_2 \cdot w_2 = \frac{2}{5} \cdot \frac{2}{3} |E| = \frac{4}{15} |E| \ .$$

Since $c_2 \cdot \chi^{M_2}$ is just $|M_2 \setminus M_1|$, it follows that

$$|M_1 \cup M_2| = \left(\frac{1}{3} + \frac{4}{15}\right) \cdot |E| = \frac{3}{5}\,|E|\,.$$

We conclude that $m_2 = 3/5$.

It remains to establish the lower bound on m_3. Note that if C is a 5-cut contained in M_1, then $w_2(C) = 1$. Hence, by Lemma 2.2, we may assume that if C is a 5-cut contained in M_1, then $|C \cap M_2| = 1$. Similarly, M_2 contains exactly one edge of each 3-cut. We now consider the following vector $w_3 \in \mathbb{R}^E$:

$$w_3(e) = \begin{cases} 1/7 & \text{if } e \in M_1 \cap M_2, \\ 2/7 & \text{if } e \in (M_1 \cup M_2) \setminus (M_1 \cap M_2), \\ 3/7 & \text{otherwise.} \end{cases}$$

Again, we verify that w_3 is a fractional perfect matching of G. The conditions (i) and (ii) hold trivially. Let us consider an odd cut C of G. If C is a 3-cut, then it contains exactly one edge e_1 contained in M_1 and exactly one edge e_2 contained if M_2. If $e_1 = e_2$, then $w_3(C) = 1/7 + 2 \cdot 3/7 = 1$. If $e_1 \neq e_2$, then $w_3(C) = 2 \cdot 2/7 + 3/7 = 1$. If C is a 5-cut that is not fully contained in M_1, then $|C \cap M_1| \leq 3$ (recall that C is an odd cut). Hence, $w_3(C) \geq 3 \cdot 1/7 + 2 \cdot 2/7 = 1$. If $C \subseteq M_1$, then $|C \cap M_2| = 1$ by the choice of M_2. We infer that $w_3(C) \geq 1/7 + 4 \cdot 2/7 > 1$. Finally, if the size of C is 7 or more, then $w_3(C) \geq 7 \cdot 1/7 = 1$. We conclude that w_3 is a fractional perfect matching of G.

Set $c_3(e) = 1 - \chi^{M_1 \cup M_2}(e)$. By Lemma 2.2, there exists a perfect matching M_3 such that

$$c_3 \cdot \chi^{M_3} = |M_3 \setminus (M_1 \cup M_2)| \geq \frac{3}{7} \cdot |E \setminus (M_1 \cup M_2)| = c_3 \cdot w_3\ .$$

Consequently,

$$\begin{aligned} |M_1 \cup M_2 \cup M_3| &= |M_1 \cup M_2| + |M_3 \setminus (M_1 \cup M_2)| \\ &\geq \frac{3}{5}\,|E| + \frac{3}{7} \cdot \frac{2}{5}\,|E| = \frac{27}{35}\,|E|\,. \end{aligned}$$

We infer that $m_3 \geq 27/35$. □

Let us remark that a similar approach as for estimating for m_2 and m_3 yields that $m_4 \geq 275/315 \approx 0.873$ (the upper bound based on Petersen graph gives $m_4 \leq 14/15 \approx 0.933$). In general, we can show that

$$m_k \geq 1 - \prod_{i=1}^{k} \frac{i+1}{2i+1}\ .$$

Acknowledgement. We would like to thank Luis Goddyn and Matt DeVos for useful and inspiring conversations on the subject.

References

[Die00] R. Diestel, *Graph Theory*, Springer, New York, 2000.

[Edm65] J. Edmonds, Maximum matching and a polyhedron with $(0,1)$-vertices, *J. Res. Nat. Bur. Standards B* **69** (1965), 125–130.

[Ful71] D. R. Fulkerson, Blocking and antiblocking pairs of polyhedra, *Math. Programming* **1** (1971), 168–194.

[SU97] E. R. Scheinerman, D. H. Ullman, *Fractional Graph Theory: A Rational Approach to the Theory of Graphs*, John Wiley & Sons, Hoboken, 1997.

[Sey79] P. D. Seymour, On multi-colourings of cubic graphs, and conjectures of Fulkerson and Tutte, *Proc. London Math. Soc.* **(3) 38** (1979), 423–460.

[Sch03] A. Schrijver, *Combinatorial Optimization*, Springer, New York, 2003.

[Zha97] C.-Q. Zhang, *Integer Flows and Cycle Covers of Graphs*, Dekker, New York, 1997.

Random Graphs from Planar and Other Addable Classes

Colin McDiarmid[1], Angelika Steger[2], and Dominic J. A. Welsh[3]

[1] University of Oxford, Department of Statistics, Oxford OX1 3TG, England
cmcd@stats.ox.ac.uk
[2] Institut für Theoretische Informatik, ETH Zürich, 8092 Zürich, Switzerland
steger@inf.ethz.ch
[3] University of Oxford, Mathematical Institute,Oxford OX1 3LB, England
dwelsh@maths.ox.ac.uk

Summary. We study various properties of a random graph R_n, drawn uniformly at random from the class $\mathcal{A}_n$ of all simple graphs on n labelled vertices that satisfy some given property, such as being planar or having tree-width at most k. In particular, we show that if the class $\mathcal{A}$ is 'small' and 'addable', then the probability that R_n is connected is bounded away from 0 and from 1. As well as connectivity we study the appearances of subgraphs, and thus also vertex degrees and the numbers of automorphisms. We see further that if $\mathcal{A}$ is 'smooth' then we can make much more precise statements for example concerning connectivity.

AMS Subject Classification. 05C80, 05C10, 68R10.
Keywords. Random graphs, planar graphs, addable graph classes.

1 Introduction

All graphs which we consider will be finite and simple. When we discuss a class $\mathcal{A}$ of graphs we shall always assume that it is closed under isomorphism. We let $\mathcal{A}_n$ denote the set of graphs in $\mathcal{A}$ on the vertex set $[n] = \{1, \ldots, n\}$. We shall only consider classes $\mathcal{A}$ such that $\mathcal{A}_n$ is non-empty for all sufficiently large n. For any non-empty finite set $\mathcal{A}$ we use the notation $R \in_u \mathcal{A}$ to mean that R is an element of $\mathcal{A}$ picked uniformly at random. Often we will consider $R_n \in_u \mathcal{A}_n$ for each n such that $\mathcal{A}_n$ is non-empty, without explicitly mentioning this condition on n.

The notion of an addable family of graphs was introduced in [MSW05], where it was used as a tool in identifying properties of random planar graphs. Although the results about addable families presented in [MSW05] can be used to derive properties for various graph classes, the emphasis in that paper was on planar graphs. Here we maintain a more general approach. Many of the results we state here are either taken from or are minor modifications of results from [MSW05]. Unless otherwise indicated, proofs of these results can be found there, either immediately or by making minor modification to the proofs given there.

2 Addable Classes of Graphs

Our main emphasis in this paper will be on classes of graphs G that have the property that adding an edge between two components of G will result in another member of the class.

Definition 2.1. *A non-empty class $\mathcal{A}$ of graphs is called* weakly addable*, if for each graph G in $\mathcal{A}$, whenever u and v are vertices in distinct components of G the graph obtained from G by adding an edge joining u and v is also in $\mathcal{A}$.*

If $\mathcal{A}$ satisfies in addition the property that a graph G belongs to $\mathcal{A}$ if and only if each component of G belongs to $\mathcal{A}$, then $\mathcal{A}$ is called addable*.*

Examples of addable graph classes include

- forests
- outerplanar graphs
- planar graphs
- series-parallel graphs
- graphs with tree-width at most k
- triangle-free graphs
- k-colorable graphs
- graphs with no cycle of length greater than k
- graphs with no K_k-minor.

Observe that if an addable class $\mathcal{A}$ contains the one vertex graphs then it must contain the class $\mathcal{F}$ of forests. An example of a class which is weakly addable but not addable, is the class of graphs embeddable on a torus. Clearly, one copy of the complete graph K_5 is embeddable on the torus, while two copies cannot be embedded simultaneously. A similar result holds for the class of graphs embeddable on any fixed orientable surface other than the sphere. The basic result for weakly addable graph classes is as follows.

Theorem 2.2. *Let $\mathcal{A}$ be any finite non-empty weakly addable class of graphs, and let $R \in_u \mathcal{A}$, so that R is a graph sampled uniformly at random from $\mathcal{A}$. Then*

- *the random number $\kappa(R)$ of components of R is stochastically dominated by $1 + X$ where $X \sim Po(1)$, that is X has the Poisson distribution with mean 1; and hence*
- $\Pr[R \text{ is connected}] \geq 1/e$, $\Pr[\kappa(R) \geq k+1] \leq 1/k!$ *and* $\mathbb{E}[\kappa(R)] \leq 2$.

By far the simplest nontrivial addable class is the class $\mathcal{F}$ of forests. Kirchhoff's well-known formula says that there are n^{n-2} connected graphs (trees) in $\mathcal{F}_n$. Rényi [Rén59], see also Moon [Moon70], showed that

$$|\mathcal{F}_n| \sim \sqrt{e} n^{n-2}. \tag{1}$$

Thus we have:

Theorem 2.3. *The probability that a random forest $R_n \in_u \mathcal{F}_n$ is connected satisfies*

$$\lim_{n\to\infty} \Pr[R_n \text{ is connected}] = \frac{1}{\sqrt{e}} \approx 0.6065.$$

[Indeed, if $R_n \in_u \mathcal{F}_n$ then $\kappa(R_n) - 1 \sim Po(\frac{1}{2})$.]

We believe that the lower bound $1/e$ from Theorem 2.2 on the probability that R_n is connected is unnecessarily low, and that in fact the lowest value is obtained for the class of forests. In other words, the class of forests is the 'least connected' weakly addable class of graphs.

Conjecture 2.4. *Let $\mathcal{A}$ be any weakly addable class of graphs. Suppose that $\mathcal{A}_n$ is non-empty for all sufficiently large n, and let $R_n \in_u \mathcal{A}_n$. Then*

$$\liminf_{n\to\infty} \Pr[R_n \text{ is connected}] \geq \frac{1}{\sqrt{e}}.$$

3 Counting

The key step in investigating random graphs from a class $\mathcal{A}$ is to count such graphs.

Definition 3.1. *A class $\mathcal{A}$ of graphs is called* small *if there exists a finite constant $a > 0$ such that we have $|\mathcal{A}_n| \leq a^n \cdot n!$ for all sufficiently large n, or equivalently if* $\limsup_{n\to\infty} (|\mathcal{A}_n|/n!)^{\frac{1}{n}} < \infty$.

Recall that, if $\mathcal{A}$ is a weakly addable class of graphs, then $\mathcal{A}$ is addable when a graph G is in $\mathcal{A}$ if and only if each component of G is in $\mathcal{A}$. The next theorem (essentially from [MSW05]) is the key result that gets things going. It applies to small addable classes of graphs, though we do not need quite the full strength of addability.

Theorem 3.2. *Let $\mathcal{A}$ be a small weakly addable class of graphs, such that if each component of a graph G is in $\mathcal{A}$ then so is G, and such that $\mathcal{A}_n$ is non-empty for all sufficiently large n. Then there is a finite constant $\gamma = \gamma(\mathcal{A}) > 0$, such that*

$$\left(\frac{|\mathcal{A}_n|}{n!}\right)^{\frac{1}{n}} \to \gamma \quad \text{as } n \to \infty.$$

When there is a finite strictly positive constant $\gamma = \gamma(\mathcal{A})$ as above, we shall call γ the *growth constant* of the class $\mathcal{A}$, and say that $\mathcal{A}$ *has a growth constant.*

The above result is stated in [MSW05] for addable classes, but the proof uses only the assumptions stated above. We may use the connectivity lower bound from Theorem 2.2 to show that $g(n) = |\mathcal{A}_n|/(e \cdot n!)$ is supermultiplicative (that is, $g(i+j) \geq g(i) \cdot g(j)$). The basic lemma on such functions then yields

$$g(n)^{1/n} \to \gamma = \sup_k g(k)^{1/k} \quad \text{as } n \to \infty.$$

We also obtain the inequality $|\mathcal{A}_n| \leq e \cdot \gamma^n \; n!$ for all n.

Let us return to the list of examples of addable classes of graphs, and investigate which are small. We have already seen that the class $\mathcal{F}$ of forests is small. The class $\mathcal{P}$ of planar graphs is small: this was proved in [DVW96] using a result of Tutte [Tut62] which shows that the number of unlabelled triangulations on n vertices grows just exponentially in n. It follows of course that the subclass of outerplanar graphs is also small. Next, the class of series-parallel graphs is small. Indeed, more generally, graphs with tree-width at most k form a small class. This follows from a surprisingly little known theorem of Beineke and Pippert (1969) and Moon (1969):

Theorem 3.3 ([BP69, Moon70]). *For each positive integer k, the number of k-trees on n vertices is given by*

$$\binom{n}{k} (k(n-k)+1)^{n-k-2}.$$

The formula shows that k-trees have growth constant ek. Since, for $n \geq k+2$, a k-tree on n vertices is constructed by adding a vertex of degree k to a k-tree on $n-1$ vertices, we know that k-trees on n vertices contain at most kn edges. Hence, also their subgraphs, the graphs of tree-width at most k, contain at most kn edges, and the graphs of tree-width at most k therefore have growth constant at most $ek2^k$.

To conclude investigating the list from the previous section, it is easy to see that triangle-free graphs and k-colorable graphs ($k \geq 2$) are not small—just consider bipartite graphs. Graphs with no cycle of length greater than k form a small class. Indeed, graphs without a K_k-minor form a small class. This follows from a recent theorem of Norine, Seymour, Thomas, and Wollan:

Theorem 3.4 ([NSTW05]). *Any proper minor-closed class $\mathcal{A}$ of graphs is small.*

As we will see, exact knowledge of the growth constant $\gamma(\mathcal{A})$ is extremely useful, but also usually quite hard to obtain. A notable early exception is the class of forests, where the value of the growth constant follows immediately from Rényi's formula (1), as it gives the much less precise result that $|\mathcal{F}_n|^{1/n} \sim \; n$: this approximation, together with the fact that $(n!)^{1/n} \sim n/e$, gives

$$\gamma(\mathcal{F}) = \lim_{n\to\infty} \left(\frac{|\mathcal{F}_n|}{n!}\right)^{1/n} = e. \tag{2}$$

Most of the other cases that are known are related to planar graphs and are based on generating function techniques which we introduce next.

4 The Generating Function Approach

If $\mathcal{A}$ is a combinatorial class and A_n denotes the number of members of $\mathcal{A}$ with size n, then the *exponential generating function* $A(x)$ is given by

$$A(x) = \sum_{n\geq 0} \frac{A_n}{n!} x^n.$$

We shall restrict our attention to graphs, with 'size' given by the number of vertices, and with the value of A_0 set appropriately depending on the context. See for example the book [Wilf90] by Wilf for an introduction to generating function methods, and the forthcoming book [FS06$^+$] by Flajolet and Sedgewick for the full story.

Observe that a class $\mathcal{A}$ of graphs is small if and only if the power series $A(x)$ converges for some real $x > 0$; and if $\mathcal{A}$ has a growth constant $\gamma = \gamma(\mathcal{A}) > 0$ then $\gamma = 1/r$, where r is the radius of convergence of $A(x)$.

One particularly nice property of the generating function method is that it relates the class of graphs (or more generally 'objects') under consideration to the class of connected graphs, and even further to the class of bi-connected graphs (or 2-vertex-connected graphs, or blocks). To emphasise this, given a class $\mathcal{A}$ of (not-necessarily-connected) graphs, we denote its exponential generating function by $A(x)$, and use $C(x)$ resp. $B(x)$ to denote the exponential generating functions of the subclass of all connected resp. bi-connected graphs in the class $\mathcal{A}$. By convention we set $|\mathcal{A}_0| = 1$, $|\mathcal{C}_0| = 0$ and $|\mathcal{B}_0| = 0$. Then, see for example [HP73], we have

Lemma 4.1. *If $\mathcal{A}$ is a class of graphs such that $G \in \mathcal{A}$ if and only if each component of G belongs to $\mathcal{A}$, then*

$$A(x) = e^{C(x)} \qquad \text{and} \qquad xC'(x) = xe^{B'(xC'(x))},$$

where $C'(x) = dC(x)/dx$ and $B'(x) = dB(x)/dx$.

The result $A(x) = e^{C(x)}$ is known as the 'exponential formula'. One way to see why this equation holds is the following. Rewrite the exponential function of the first equation as a power series to obtain $\sum_{k\geq 0} C(x)^k/k!$. The coefficient of x^n on the left hand side $A(x)$ is, by definition, the number of graphs $|\mathcal{A}_n|$ divided by $n!$. The coefficient of x^n in $C(x)^k/k!$ is

$$\sum_{(n_1,\dots,n_k)} \frac{1}{k!} \frac{|C_{n_1}|}{n_1!} \cdot \dots \cdot \frac{|C_{n_k}|}{n_k!} = \frac{1}{n!} \cdot \frac{1}{k!} \sum_{(n_1,\dots,n_k)} \binom{n}{n_1,\dots,n_k} |C_{n_1}| \cdot \dots \cdot |C_{n_k}|,$$

where the sums are over all *ordered* tuples $(n_1, \ldots, n_k)$ of positive integers which sum to n. One easily sees that the right hand side here is the number of graphs in $\mathcal{A}_n$ which consist of exactly k components, divided by $n!$. Summing over k we thus find that for all n the coefficients of x^n on the left and right hand side of $A(x) = e^{C(x)}$ are the same.

For the second equation in Lemma 4.1 one may argue similarly. We give here an even briefer sketch. First we observe that the left side $xC'(x)$ counts the number graphs in $\mathcal{C}_n$ that are *rooted* at an (arbitrary) vertex. Now the blocks (2-vertex-connected components) of a connected graph form a tree. If we root that tree at an arbitrary vertex r then r will be in some number k of blocks, which are each rooted at r. The corresponding sum over k is taken care of by the exponential function, as in the first equation. The argument of the exponential is given by the function B', as we are looking at rooted blocks. Its argument is $xC'(x)$ instead of simply x, as each vertex (except the root r) of the blocks containing r can again be the root of a connected component.

Consider a class $\mathcal{A}$ of graphs such that G is in $\mathcal{A}$ if and only if each component of G is in $\mathcal{A}$. Because of the relation $A(x) = e^{C(x)}$ we see that the two power series $A(x)$ and $C(x)$ must have the same radius of convergence. Thus $\mathcal{A}$ is small if and only if $\mathcal{C}$ is small. Also, if $\mathcal{C}$ has a growth constant then so does $\mathcal{A}$ and the growth constants are the same. For the class of forests this immediately gives another proof of the fact that $\gamma(\mathcal{F}) = e$.

Let us also note that the converse of the last implication is false in general. Consider, for example, the class $\mathcal{A}$ of all forests such that each component is either a single vertex or has an even number of vertices. Then $\mathcal{A}$ satisfies the condition that G is in $\mathcal{A}$ if and only if each component of G is in $\mathcal{A}$, and $\mathcal{A}$ has growth constant $\gamma(\mathcal{A}) = e$, but the class of connected graphs in $\mathcal{A}$ does not have a growth constant.

Using Lemma 4.1 together with the important asymptotic estimates for bi-connected planar graphs by Bender, Gao, and Wormald [BGW02], Giménez and Noy were able to obtain the following remarkable complete picture for planar graphs:

Theorem 4.2 ([GN05]). *Let g_n be the number of planar graphs on n vertices. Then*

$$g_n \sim g \cdot n^{-7/2} \gamma^n n!, \tag{3}$$

where $g \approx 4.26094 \cdot 10^{-6}$ and $\gamma \approx 27.2269$ are constants given by explicit analytic expressions. Furthermore, let c_n be the number of connected planar graphs on n vertices. Then

$$c_n \sim c \cdot n^{-7/2} \gamma^n n!,$$

where γ is as before and $c \approx 4.10436 \cdot 10^{-6}$.

Giménez and Noy [GN05] also show related results concerning the random planar graph R_n. In particular, they show that its number of edges is asymptotically normally distributed with mean $\sim \kappa n$ and variance $\sim \lambda n$, where

$\kappa \approx 2.2133$ and $\lambda \approx 0.4304$. They also give sharp asymptotic results on the connectivity and the number of components, see Theorem 9.3 below.

Using similar techniques the growth constants were recently also determined for various related classes of graphs: for cubic planar graphs [BLKM05] the growth constant is 3.1326, and for outerplanar and series parallel graphs [BGKN06$^+$] the growth constants are 7.3210 and 9.0733 respectively (all to four decimal places). For the cubic case we must naturally restrict our attention to even numbers of nodes. We do not know for example about 4-regular or 5-regular planar graphs.

5 The Appearances Theorem

Suppose that $\mathcal{A}$ is a small addable class of graphs, let $R_n \in_u \mathcal{A}_n$, and let H be a fixed connected graph in $\mathcal{A}$. Then with high probability, R_n contains linearly many copies of H. Let us be more precise, and more general. We need some definitions.

Let H be a graph on the vertex set $\{1, \ldots, h\}$, and let G be a graph on the vertex set $\{1, \ldots, n\}$ where $n > h$. Let $W \subset V(G)$ with $|W| = h$, and let the 'root' r_W denote the least element in W. We say that H *appears* at W in G if (a) the increasing bijection from $\{1, \ldots, h\}$ to W gives an isomorphism between H and the induced subgraph $G[W]$ of G; and (b) there is exactly one edge in G between W and the rest of G, and this edge is incident with the root r_W. We let $f_H(G)$ be the number of appearances of H in G, that is the number of sets $W \subseteq V(G)$ such that H appears at W in G.

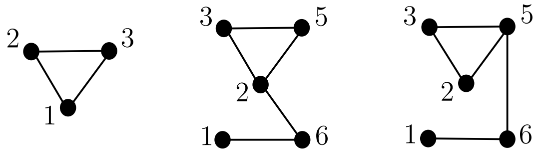


Fig. 1. A graph H, an appearance of H, and a non-appearance of H

Let $\mathcal{A}$ be a class of graphs, and let H be a fixed graph with a distinguished 'root' vertex r. We say that H can be *attached* to $\mathcal{A}$ if, for each graph G in $\mathcal{A}$ and each vertex v in G, the graph obtained as follows must be in $\mathcal{A}$: take vertex disjoint copies of G and H and make the root r of H adjacent to vertex v in G.

Theorem 5.1. *Let the class $\mathcal{A}$ of graphs have a (finite positive) growth constant, and let $R_n \in_u \mathcal{A}_n$. Let H be a fixed graph with a distinguished 'root' vertex r, and suppose that H can be* attached *to $\mathcal{A}$ in the above sense. Then there exists a constant $\alpha > 0$ such that*

$$\Pr[f_H(R_n) \leq \alpha n] = e^{-\Omega(n)}. \tag{4}$$

The bound $e^{-\Omega(n)}$ in equation 4 above means that there exists a constant $\beta > 0$ (depending on $\mathcal{A}$ and H) such that the probability is at most $e^{-\beta n}$ for all sufficiently large n. Note that by Theorem 3.2 the conditions of the above theorem must hold if $\mathcal{A}$ is small and addable and contains H, and $\mathcal{A}_n$ is non-empty for sufficiently large n. The way we formulated the theorem, however, allows it to be applied to non-addable graph classes. Consider, for example, the class $\mathcal{A}$ of all graphs embeddable on a torus. Then $\mathcal{A}$ is not addable, but nevertheless we can attach every planar graph H to $\mathcal{A}$ in the sense above.

Theorem 5.1 may be proved along the lines of the proof of Theorem 4.1 in [MSW05]. The main idea is to show that if the probability in (4) is not very small for n vertices then one can construct "too many" graphs on $(1+\delta)n$ vertices. Essentially this is done by attaching linearly many copies of H to each graph G on n vertices which has "few" appearances of H: this property of G limits the amount of double-counting involved.

Let us give two applications of Theorem 5.1, first to vertex degrees, and then to automorphisms and unlabelled graphs.

5.1 Vertex Degrees

By applying Theorem 5.1 to especially chosen graphs H, for example to a star on k vertices, we can deduce from it some results about the vertex degrees in a random graph R_n from $\mathcal{A}_n$.

Corollary 5.2. *Let the class $\mathcal{A}$ of graphs have a growth constant, and let $R_n \in_u \mathcal{A}_n$. Let $k \geq 1$ and suppose that the star with $k-1$ edges, rooted at its centre, can be attached to $\mathcal{A}$. Then there exists a constant $\alpha > 0$ such that*

$$\Pr[R_n \textit{ has } < \alpha n \textit{ vertices of degree } k\ ] = e^{-\Omega(n)}.$$

Thus we have many vertices of each fixed degree in for example planar graphs and graphs of bounded tree-width. The next result tells us about vertices of high degree.

Corollary 5.3. *Let the class $\mathcal{A}$ of graphs have a growth constant, and be closed under adding a pendant vertex or deleting a pendant edge. Then with probability tending to 1 as $n \to \infty$, the random graph $R_n \in_u \mathcal{A}_n$ has a vertex of degree at least $(1+o(1))(\ln n/\ln\ln n)$ (and indeed R_n has a vertex adjacent to this number of pendant vertices).*

We spell out a proof here, as this was not done in [MSW05]. The main idea is to partition the set $\mathcal{A}_n$ into suitable subclasses such that we can show that within each subclass most graphs contain a vertex of large degree.

For each graph $G \in \mathcal{A}$ let $W(G)$ be the set of vertices of degree 1 which are adjacent to a vertex in a triangle, and let $H(G)$ be the graph induced on the vertices not in $W(G)$. Let $\mathcal{A}_n(W,H)$ be the set of graphs $G \in \mathcal{A}_n$ such that $W(G) = W$ and $H(G) = H$. Then $\mathcal{A}_n$ is partitioned into the non-empty sets of the form $\mathcal{A}_n(W,H)$.

Suppose that we condition on $W(R_n) = W$ and $H(R_n) = H$, where H contains exactly n' vertices in triangles. Then the maximum number of edges in R_n between a vertex in H and the vertices in W has the same distribution as the maximum bin load when we throw $|W|$ balls independently at random into n' bins. By Theorem 5.1 there exists $\alpha > 0$ such that $|W(R_n)| \geq \alpha n$ with probability tending to 1 as $n \to \infty$. But a standard result concerning balls and bins shows that if we throw at least αn balls into at most n bins, then the maximum bin load is $(1 + o(1)) \ln n / \ln \ln n$ with high probability, and Corollary 5.3 follows.

Can this last result be improved? For a random tree T_n on n vertices, Moon [Moon68] shows that

$$\Pr[\Delta(T_n) \geq d + 1] \leq \frac{n}{d!} \qquad \text{for all } d \geq 3.$$

From this one may easily deduce that a random forest $F_n \in_u \mathcal{F}_n$ similarly satisfies

$$\Pr[\Delta(F_n) \geq d + 1] \leq \frac{n}{d!} \qquad \text{for all } d \geq 3,$$

by partitioning the vertex sets into its components. It now follows easily that $\Delta(F_n) \leq (1 + o(1)) \ln n / \ln \ln n$ with probability $\to 1$ as $n \to \infty$, which shows that the bound in Corollary 5.3 is essentially best possible for the class of forests.

Question. Is this true more generally? In particular is it true for a random planar graph R_n that the maximum degree $\Delta(R_n)$ is $(1 + o(1)) \ln n / \ln \ln n$ with high probability?

5.2 Automorphisms and Unlabelled Graphs

We continue with corollaries of the appearances theorem, Theorem 5.1. The next result follows immediately from the fact that a graph with linearly many vertices that are connected to at least two vertices of degree one has exponentially many automorphisms.

Corollary 5.4. *Let the class $\mathcal{A}$ of graphs have a growth constant, and suppose that the 3-vertex path, rooted at its centre, can be attached to $\mathcal{A}$ (as defined preceding Theorem 5.1). Then there exists a constant $\alpha > 0$ such that $R_n \in_u \mathcal{A}_n$ satisfies*

$$\Pr[R_n \textit{ has } < 2^{\alpha n} \textit{ automorphisms }] = e^{-\Omega(n)}.$$

This result is in contrast to what happens with classical random graphs: for example, $G_{n,\frac{1}{2}}$ has only the trivial automorphism with probability $\to 1$ as $n \to \infty$. Note that we could replace the 3-vertex path above by any rooted graph with a non-trivial automorphism fixing its root.

The *unlabelled* graphs in $\mathcal{A}$ are just the automorphism classes of the (labelled) graphs in $\mathcal{A}$. Let $\mathcal{UA}$ denote the set of unlabelled graphs in $\mathcal{A}$, and let $\mathcal{UA}_n$ be the set of unlabelled graphs in $\mathcal{A}$ with n vertices.

Recall that for the class $\mathcal{F}$ of (labelled) forests we have $(|\mathcal{F}_n|/n!)^{1/n} \to \gamma(\mathcal{F}) = e$ as $n \to \infty$. Now let us consider unlabelled trees and forests. A classical theorem of Otter [Ott48] states that the number t_n of unlabelled trees on n vertices satisfies

$$t_n \sim c \cdot n^{-5/2} \eta^n,$$

where $c \approx 0.5349$ and $\eta \approx 2.9557$. Palmer and Schwenk [PS79] study the number of unlabelled trees in a random forest. In particular, they show that the probability that a random forest consists of just one tree, in other words that the forest is connected, converges to a constant ≈ 0.5228. Thus the number $|\mathcal{UF}_n|$ of unlabelled forests on n vertices satisfies $|\mathcal{UF}_n|^{1/n} \to \gamma(\mathcal{UF}) = \eta$. (For unlabelled graphs we do not divide by $n!$.) Note that $\gamma(\mathcal{UF}) = \eta > e = \gamma(\mathcal{F})$. There is a corresponding general result.

Corollary 5.5. *Let the class $\mathcal{A}$ of graphs have growth constant γ, and suppose that the 3-vertex path, rooted at its centre, can be attached to $\mathcal{A}$. Then there exists $\tilde{\gamma} > \gamma$ such that the number $|\mathcal{UA}_n|$ of unlabelled graphs in $\mathcal{A}_n$ is at least $\tilde{\gamma}^n$ for all sufficiently large n.*

This result may be proved along the lines of the proof of Corollary 4.7 in [MSW05]. The isomorphism class of a graph $G \in \mathcal{A}_n$ has size $n!/aut(G)$, where $aut(G)$ is the size of the automorphism group of G. By Corollary 5.4, the number of graphs $G \in \mathcal{A}_n$ which are in automorphism classes of size $> 2^{-\alpha n} n!$ is at most $\frac{1}{2}|\mathcal{A}_n|$ for all sufficiently large n. But then

$$|\mathcal{UA}_n| \geq \frac{1}{2}|\mathcal{A}_n|/(2^{-\alpha n} n!) = \frac{1}{2} 2^{\alpha n} |\mathcal{A}_n|/n! = (2^{\alpha}\gamma + o(1))^n.$$

Some questions on unlabelled graphs. Consider a class $\mathcal{A}$ of (labelled) graphs with a growth constant. Must there exist β such that the number $|\mathcal{UA}_n|$ of unlabelled graphs in $\mathcal{A}_n$ is at most β^n for all sufficiently large n? When will $|\mathcal{UA}_n|^{1/n}$ tend to a limit as $n \to \infty$? It is known for example that for the class $\mathcal{UP}$ of unlabelled planar graphs we have $|\mathcal{UP}_n|^{1/n}$ tends to a limit $\gamma(\mathcal{UP})$: what is this constant? It is known how to generate $R_n \in_u \mathcal{P}_n$ in time bounded polynomially in n [BGK03], [Fusy06$^+$]. Can we similarly generate an unlabelled graph from $\mathcal{UP}_n$ uniformly at random in time bounded polynomially in n?

6 Connectivity and Components

We have seen in Theorem 2.2 that, for a range of graph classes, the probability that a randomly chosen member of the class is connected can be bounded away from zero. This probability can also be bounded away from one. The following result is based on Theorem 5.1 of [MSW05].

Theorem 6.1. *Let the class $\mathcal{A}$ of graphs have a growth constant, and assume that $\mathcal{A}$ is weakly addable and is closed under adding an isolated vertex and deleting a pendant edge. Then $R_n \in_u \mathcal{A}_n$ satisfies*

$$\liminf_{n\to\infty} \Pr[R_n \textit{ has an isolated vertex}] > 0.$$

The key step in proving this is to use Theorem 2.2 and Corollary 5.2 to show that with probability at least $1/e + o(1)$, R_n is both connected and has at least αn vertices of degree 1. Now deleting an edge incident with a vertex of degree 1 yields a graph with exactly one isolated vertex. We obtain many such graphs in this way, indeed at least $(1/e + o(1))|\mathcal{A}_n| \cdot \alpha n$, and each graph is produced at most n times. Thus the probability that R_n has an isolated vertex is at least $\alpha/e + o(1)$.

A similar result holds if we replace the isolated vertex in this last result by any connected graph H.

Theorem 6.2. *Let the class $\mathcal{A}$ of graphs have a growth constant, and assume that $\mathcal{A}$ is weakly addable and is closed under deleting bridges. Let H be a (fixed) connected graph, and suppose that for any $G \in \mathcal{A}$ the union of G with a disjoint copy of H is also in $\mathcal{A}$. Then $R_n \in_u \mathcal{A}_n$ satisfies*

$$\liminf_{n\to\infty} \Pr[R_n \textit{ contains a component isomorphic to } H] > 0.$$

Theorem 2.2 showed that, for a weakly addable class of graphs, we should expect few components. There is a version of Theorem 2.2, with essentially the same proof, showing that large components are even less likely to appear, and which will allow us to see that there is usually a 'giant component'. In order to state this version we need one more piece of notation. Given a positive integer k, let $\kappa_k(G)$ be the number of components of the graph G with order (number of vertices) at least k.

Theorem 6.3 ([MSW05]). *Let $\mathcal{A}$ be any finite non-empty weakly addable class of graphs, and let $R \in_u \mathcal{A}$. Let k be a positive integer. Then $\kappa_k(R)$ is stochastically dominated by $1 + X$ where $X \sim Po(1/k)$, and hence*

$$\Pr[R \textit{ has more than one component with order} \geq k] \leq 1 - e^{-1/k} \leq 1/k.$$

Now for the 'giant component'.

Theorem 6.4. *Let $\mathcal{A}$ be any weakly addable class of graphs such that $\mathcal{A}_n$ is non-empty for all sufficiently large n. Then for any $\varepsilon > 0$ there exists a constant $k = k(\varepsilon)$ such that, for all sufficiently large n, the probability is at least $1 - \varepsilon$ that $R_n \in_u \mathcal{A}_n$ has a component (the* giant *component) which contains all but at most k vertices.*

To prove this, let a and b be positive integers and use Theorems 2.2 and 6.3 to see that the probability that R_n has no component containing all but at most ab vertices is at most

$$\Pr[\kappa_{a+1}(R_n) \geq 2] + \Pr[\kappa(R_n) \geq b + 1] \leq 1/(a + 1) + 1/b!.$$

7 Expected Numbers of Vertices of Low Degree

Let $\mathcal{A}$ be a class of graphs, suppose that $\mathcal{A}_n$ is non-empty, and let $R_n \in_u \mathcal{A}_n$. Let $X_k = X_k(R_n)$ denote the number of vertices in R_n of degree k. We are interested here in $\mathbb{E}[X_k]$ for $k = 0$ in particular, in other words the expected number of isolated vertices, but also for $k = 1$ and 2.

Suppose first that $\mathcal{A}$ is closed under adding or deleting an isolated vertex. For example we could consider forests or planar graphs. Then it is clear that

$$\Pr[\text{ vertex } n \text{ is isolated in } R_n\,] = \frac{|\mathcal{A}_{n-1}|}{|\mathcal{A}_n|},$$

and so, as $\mathcal{A}$ is closed under isomorphisms, we have

$$\mathbb{E}[X_0] = \frac{n|\mathcal{A}_{n-1}|}{|\mathcal{A}_n|}. \tag{5}$$

For example, by equation (1), for forests the expected number of isolated vertices tends to e^{-1}.

Now assume that $\mathcal{A}$ is closed under adding or deleting a pendant vertex. Then arguing as above, we obtain

$$\mathbb{E}[X_1] = \frac{n(n-1)|\mathcal{A}_{n-1}|}{|\mathcal{A}_n|}. \tag{6}$$

Thus $\mathbb{E}[X_1] = (n-1)\mathbb{E}[X_0]$, if $\mathcal{A}$ is also closed under adding or deleting an isolated vertex. Hence, from the above, for forests the expected number of pendant vertices $\sim e^{-1}n$. This follows also from equations (1) and (6), and indeed we find the same result for trees.

Finally consider the following condition on the class $\mathcal{A}$ of graphs, that for each graph $G \in \mathcal{A}$ and each pair of distinct vertices u and v in G:

(a) if either uv is an edge, or uv is not an edge and $G + uv \in \mathcal{A}$, then the graph obtained by adding a new degree 2 vertex adjacent to u and v is in $\mathcal{A}$; and
(b) if there is a vertex x of degree 2 adjacent to u and v then the graph G' obtained by deleting x (and the two incident edges) is in $\mathcal{A}$, and further if uv is not an edge then $G' + uv \in \mathcal{A}$.

For example, planar graphs (and more generally graphs embeddable on any given surface) and graphs of tree width at most k (for $k \geq 2$) satisfy this condition. For a graph class $\mathcal{A}$ satisfying this condition, we have

$$\mathbb{E}[X_2] = \mathbb{E}[X_1] \cdot \bar{d}(n-1), \tag{7}$$

where $\bar{d}(n-1)$ denotes the average degree of R_{n-1} (see the proof of Theorem 4.10 of [MSW05]).

Equations (5), (6) and (7) each apply to the class of planar graphs. Hence, from the results of [GN05] discussed at the end of Section 4, we see that for planar graphs, $\mathbb{E}[X_0] \to \gamma^{-1} \approx 0.0367$, $\mathbb{E}[X_1] \sim \gamma^{-1}n \approx 0.0367n$ and $\mathbb{E}[X_2] \sim 2\gamma^{-1}\kappa n \approx 0.1626n$; and so the expected number of vertices of degree at least 3 is $\sim (1-\gamma^{-1}(1+2\kappa))n \approx 0.8007n$.

8 Smoothness

Recall from Section 4 that a class $\mathcal{A}$ of graphs has exponential generating function $A(x) = \sum_{n\geq 0} \frac{|\mathcal{A}_n|}{n!} x^n$. Denote $\frac{|\mathcal{A}_n|}{n!}$ by t_n, so $A(x)$ is the power series $\sum_{n\geq 0} t_n x^n$, where we may think of x as a real (or complex) variable. The *exponential growth rate* of $\mathcal{A}$ is defined [FS06+] to be $\limsup_{n\to\infty} t_n^{1/n} = \gamma$, say, where $0 \leq \gamma \leq \infty$. Note that the radius of convergence of $A(x)$ is γ^{-1}. We said earlier that $\mathcal{A}$ is small when $0 < \gamma < \infty$; and that $\mathcal{A}$ has growth constant γ when $t_n^{1/n} \to \gamma$ as $n \to \infty$. Now consider the ratio of successive coefficients of $A(x)$, namely

$$\frac{t_n}{t_{n-1}} = \frac{|\mathcal{A}_n|}{n|\mathcal{A}_{n-1}|}.$$

It is well known and easy to see that if as $n \to \infty$, $\frac{t_n}{t_{n-1}}$ tends to a limit then $t_n^{1/n}$ tends to the same limit. (This is the basis of the ratio method for finding the radius of convergence of a power series.) Thus if $\frac{t_n}{t_{n-1}}$ tends to a limit γ, then $\mathcal{A}$ has growth constant γ. We call $\mathcal{A}$ *smooth* in this case, that is, when the ratios $\frac{t_n}{t_{n-1}}$ tend to a limit as $n \to \infty$.

It was conjectured in [MSW05] that the class of planar graphs was smooth. Observe that by the equation (5) this is equivalent to conjecturing that the expected number of isolated vertices tends to a limit as $n \to \infty$ – thus the name *isolated vertices conjecture* used in [MSW05]. Precise results were given in that paper on for example the limiting probability of connectedness, on the assumption that the conjecture was true. In fact, smoothness follows immediately from the recent asymptotic formula (3) of Giménez and Noy [GN05].

Proving smoothness for a class $\mathcal{A}$ of graphs seems to be difficult, unless we know enough about the exponential generating function $A(x)$ to derive an asymptotic formula for $|\mathcal{A}_n|$, like (3) for planar graphs. But as soon as we have a suitable explicit asymptotic formula, smoothness follows immediately. From results mentioned earlier we thus know smoothness for the classes of forests, planar graphs, outerplanar graphs, and series-parallel graphs. (For cubic planar graphs we need to adapt the definitions of growth constant and smoothness, since such graphs must have an even number of vertices.)

There are examples of graph classes which have a growth constant and are reasonably well behaved, but which are not smooth. For example, consider the class of graphs G such that G less any isolated vertices is cubic and planar:

this class has a growth constant but is not smooth. However, it seems that any 'natural' small addable class $\mathcal{A}$ of graphs is smooth. To formulate a conjecture, let us call a class $\mathcal{A}$ of graphs *hereditary* if whenever a graph G is in $\mathcal{A}$ then so is each subgraph of G (not just induced subgraphs). We make the following *Smoothness Conjecture*:

Conjecture 8.1. *Any non-empty small addable hereditary class $\mathcal{A}$ of graphs is smooth.*

In the next section we shall demonstrate the importance of the concept of smoothness, by showing that smoothness yields results on connectivity and connectedness which are much more precise than those in Section 6.

9 Connectivity and Components for Smooth Classes

In this final section we explore properties of a random graph $R_n \in_u \mathcal{A}_n$ for a smooth class $\mathcal{A}$ of graphs. Perhaps quite surprisingly, smoothness allows us to relate the probability of connectedness to the growth constant. The results in this section are based on section 5 of [MSW05].

We first consider the probability that R_n contains an isolated vertex. In this case the formula will turn out to be particular easy.

Theorem 9.1. *Let $\mathcal{A}$ be a smooth class of graphs, with growth constant $\gamma = \gamma(\mathcal{A})$, which is closed under adding an isolated vertex. Then the number of isolated vertices in $R_n \in_u \mathcal{A}_n$ is asymptotically Poisson distributed with parameter $1/\gamma$; and so*

$$\Pr[R_n \text{ contains an isolated vertex}] \;\to\; 1 - e^{-1/\gamma} \quad \text{as } n \to \infty.$$

The above result shows how smoothness lets us strengthen Theorem 6.1. The next, more general, result does a similar job for Theorem 6.2.

Theorem 9.2. *Let $\mathcal{A}$ be a smooth class of graphs, with growth constant $\gamma = \gamma(\mathcal{A})$. Let H be a fixed connected graph such that, for each $G \in \mathcal{A}$, the union of G with a disjoint copy of H is also in $\mathcal{A}$. Then the number of components isomorphic to H in $R_n \in_u \mathcal{A}_n$ is asymptotically Poisson distributed with parameter $\lambda := 1/(aut(H)\gamma^{|V(H)|})$; and so*

$$\Pr[R_n \text{ contains a component isomorphic to } H] \;\to\; 1 - e^{-\lambda} \quad \text{as } n \to \infty.$$

There is an even more general version of these last results, involving several non-isomorphic connected graphs $H_1, \ldots, H_k$ simultaneously. This result, together with Theorem 6.4 on the giant component, allows us to compute the probability that a random graph is connected. The following result is essentially a rewriting of the statement of Theorem 5.6 of [MSW05] in terms of exponential generating functions. Recall that $\kappa(G)$ denotes the number of components of G.

Theorem 9.3. *Let $\mathcal{A}$ be an addable smooth class of graphs, with growth constant $\gamma = \gamma(\mathcal{A})$. Then $\kappa(R_n) - 1$ is asymptotically Poisson distributed with parameter $\lambda := C(1/\gamma)$, where $C(x)$ is the exponential generating function of the connected members of $\mathcal{A}$; and so in particular*

$$\Pr[R_n \textit{ is connected}] \rightarrow 1 - e^{-\lambda} \quad \textit{as } n \rightarrow \infty.$$

For a version of this result for planar graphs see Theorem 6 and Corollary 1 of [GN05], where the corresponding value of λ is given as 0.037439, and so $\Pr[R_n$ is connected$] \rightarrow 0.963253$ (figures to six decimal places). For series-parallel and outerplanar graphs, the corresponding values of λ are 0.117614 and 0.148404 [BGKN06+].

Acknowledgement. We would like to acknowledge helpful comments from Chris Dowden, Stefanie Gerke and Anna de Mier.

References

[BP69] L. W. Beineke and R. E. Pippert, The number of labelled k-dimensional trees, *Journal of Combinatorial Theory* **6** (1969) 200–205.

[BGW02] A. Bender, Z. Gao and N. Wormald, The number of labeled 2-connected planar graphs, *Electronic Journal of Combinatorics* **9** (2002) #R43.

[BGKN06+] M. Bodirsky, O. Giménez, M. Kang and M. Noy, On the number of series-parallel and outerplanar graphs, in *Proceedings of European Conference on Combinatorics, Graph Theory, and Applications (EuroComb 2005)*, accepted for publication.

[BLKM05] M. Bodirsky, M. Löffler, M. Kang and C. McDiarmid, Random cubic planar graphs, manuscript, 2005.

[BGK03] M. Bodirsky, C. Gröpl and M. Kang, Generating labeled planar graphs uniformly at random, ICALP 2003, Springer LNCS 2719 (2003) 1095–1107.

[DVW96] A. Denise, M. Vasconcellos and D. Welsh, The random planar graph, *Congressus numerantium* **113** (1996) 61–79.

[FS06+] P. Flajolet, and R. Sedgewick, *Analytic Combinatorics* (book in preparation), preliminary version available at `http://algo.inria.fr/flajolet/Publications`.

[Fusy06+] É. Fusy, Quadratic and linear time random generation of planar graphs (extended abstract) to appear in *Discrete Mathematics and Theoretical Computer Science*

[GN05] O. Giménez and M. Noy, Asymptotic enumeration and limit laws of planar graphs, arXiv: math.CO/0501269, 2005.

[HP73] F. Harary and E. Palmer, *Graphical Enumeration*, Academic Press, New York – London, 1973.

[MSW05] C. McDiarmid, A. Steger, and D. Welsh, Random planar graphs, *Journal of Combinatorial Theory, Series B* **93**, 2005,187–205.

[Moon70] J. W. Moon, *Counting Labelled Trees*, Canadian Mathematical Monographs No 1, 1970.

[Moon68] J. W. Moon, On the maximum degree in a random tree, *Michigan Math. Journal* **15**, 1968, 429–432.

[NSTW05] S. Norine, P. Seymour, R. Thomas, and P. Wollan, Proper minor-closed families are small, manuscript, 2005.

[Ott48] R. Otter, The number of trees, *Ann. of Math.* **49** (1948) 583–599.

[PS79] E. M. Palmer and A. J. Schwenk, On the number of trees in a random forest, *Journal of Combinatorial Theory, Series B* **27** (1979) 109–121.

[Rén59] A. Rényi, Some remarks on the theory of trees, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **4** (1959) 73–85.

[Tut62] W. T. Tutte, A census of planar triangulations, *Canadian Journal of Mathematics* **14** (1962), 21–38.

[Wilf90] H. S. Wilf, *generatingfunctionology*, Academic Press, 1990.

Extremal Hypergraph Problems and the Regularity Method

Brendan Nagle[1]*, Vojtěch Rödl[2]**, and Mathias Schacht[3]***

[1] Department of Mathematics and Statistics, University of Nevada, Reno, NV 89557, USA
nagle@unr.edu
[2] Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA
rodl@mathcs.emory.edu
[3] Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099, Berlin, Germany
schacht@informatik.hu-berlin.de

Summary. Szemerédi's regularity lemma asserts that every graph can be decomposed into relatively few random-like subgraphs. This random-like behavior enables one to find and enumerate subgraphs of a given isomorphism type, yielding the so-called counting lemma for graphs. The combined application of these two lemmas is known as the *regularity method for graphs* and has proved useful in graph theory, combinatorial geometry, combinatorial number theory and theoretical computer science.

Recently, the graph regularity method was extended to hypergraphs by Gowers and by Skokan and the authors. The *hypergraph regularity method* has been successfully employed in a handful of combinatorial applications, including alternative proofs to well-known density theorems of Szemerédi and of Furstenberg and Katznelson. In this paper, we apply the hypergraph regularity method to a few extremal hypergraph problems of Ramsey and Turán flavor.

AMS Subject Classification. Primary: 05C35. Secondary: 05C65, 05D10.
Keywords. Turán's theorem, Ramsey theory, removal lemma, regularity lemma for hypergraphs

1 Introduction

Szemerédi's regularity lemma asserts that every graph can be decomposed into a bounded number of so-called ε-regular pairs. For a graph $G = (V, E)$

* Supported by NSF grant DMS 0501090.
** Supported by NSF grant DMS 0300529.
*** Supported by DFG grant SCHA 1263/1-1.

and $\varepsilon > 0$, we say two non-empty disjoint subsets $X, Y \subset V$ are *ε-regular* if for all $X' \subseteq X$, $|X'| > \varepsilon|X|$ and $Y' \subseteq Y$, $|Y'| > \varepsilon|Y|$, we have $|d_G(X,Y) - d_G(X',Y')| < \varepsilon$, where $d_G(X',Y') = |G[X',Y']|/(|X'||Y'|)$ is the *density* of the bipartite subgraph $G[X',Y']$ of G (consisting of all edges $\{x,y\} \in E$ with $x \in X'$ and $y \in Y'$). Szemerédi's lemma is then given as follows.

Theorem 1.1 (Szemerédi's regularity lemma). *For every $\varepsilon > 0$ and integer t_0, there exist integers $T_0 = T_0(\varepsilon, t_0)$ and $N_0 = N_0(\varepsilon, t_0)$ so that for every graph $G = (V, E)$, $|V| \geq N_0$, V admits a partition $V = V_1 \cup \ldots \cup V_t$, $t_0 \leq t \leq T_0$, satisfying*

(i) $|V_1| \leq \ldots \leq |V_t| \leq |V_1| + 1$ *and*
(ii) all but at most $\varepsilon\binom{t}{2}$ pairs (V_i, V_j), $1 \leq i < j \leq t$, are ε-regular.

Partitions $V = V_1 \cup \ldots \cup V_t$ satisfying (i) and (ii) as above are said to be *t-equitable* and *ε-regular*. Szemerédi's regularity lemma lead to many applications in combinatorial mathematics, particularly in the area of extremal graph theory (see [KS96, KSSS02] for surveys). Many applications of Szemerédi's lemma depend on the fact that within an appropriately given ε-regular partition $V = V_1 \cup \ldots \cup V_t$, one may enumerate small subgraphs of a fixed isomorphism type. This result is formally due to the following easily proved 'counting lemma' for graphs. In Fact 1.2 below and elsewhere in this paper, we write $x = y \pm \xi$ for reals x and y and some positive $\xi > 0$ for the inequalities $y - \xi < x < y + \xi$.

Fact 1.2 (Graph counting lemma). *For all $d > 0$, $\gamma > 0$ and every positive integer ℓ, there exist $\varepsilon > 0$ and n_0 so that whenever G is an ℓ-partite graph with ℓ-partition $V_1 \cup \ldots \cup V_\ell$, and $|V_1| = \ldots = |V_\ell| = n \geq n_0$, satisfying for all $1 \leq i < j \leq \ell$*

(a) $d_G(V_i, V_j) = d \pm \varepsilon$ *and*
(b) (V_i, V_j) *is ε-regular,*

then the number $|\mathcal{K}_\ell(G)|$ of ℓ-cliques in G satisfies $|\mathcal{K}_\ell(G)| = d^{\binom{\ell}{2}} n^\ell (1 \pm \gamma)$.

We refer to a joint application of Theorem 1.1 and Fact 1.2 as the *regularity method* for graphs. Perhaps one of the first applications of this method is due to Ruzsa and Szemerédi [RS78] who showed it can be used to prove Roth's theorem [Roth52, Roth53], i.e., Theorem 1.4 below for $d = 1$ and $\ell = 3$. More formally, Ruzsa and Szemerédi used the graph regularity method to prove that every graph G_n on n vertices having $o(n^3)$ triangles contains a triangle-free subgraph G_n' having only $o(n^2)$ fewer edges. Their result can be referred to as the 'triangle removal lemma' (cf. Theorem 1.3 below) and implies, as a corollary, Roth's theorem.

In what follows, a *hypergraph* $\mathcal{H} \subseteq 2^V$ with vertex set V is a collection of subsets from V. We say $\mathcal{H}^{(k)}$ is a *k-uniform hypergraph*, or *k-graph*, for short, if every subset belonging to $\mathcal{H}^{(k)}$ has cardinality k.

An extension of Szemerédi's regularity lemma for 3-graphs has been developed in [FR02]. More recently, extensions to k-graphs were obtained by Gowers [Gow06$^+$, Gow06] and by Skokan and the current authors [NRS06, RS04], and based on that work, subsequently by Tao [Tao06$^+$b] and the second two authors [RS06$^+$]. Using these techniques, a handful of 3-graph applications appear in [ARS06$^+$, FR02, HNR03, KNR02, KNR03, NR98, RR98, RRS06, SS05, Sol03] and some applications for k-graphs appear in [NRS06, RSST06$^+$, RSTT06$^+$, RS06$^+$] (some of which we discuss momentarily). Tao also obtained some deep number-theoretic applications in [Tao06$^+$a].

The goal of this paper is to use the hypergraph regularity method established in [NRS06, RS04] to investigate some extremal hypergraph problems (see Section 2). The components of the hypergraph regularity method, i.e., the hypergraph regularity lemma of [RS04] and the hypergraph counting lemma of [NRS06], are technical statements which we will only present later in Section 4 (cf. Remark 2.5). The following so-called *removal lemma*, however, is a direct consequence of the regularity method for hypergraphs.

Theorem 1.3 (Removal lemma, [Gow06$^+$, NRS06, RS06$^+$]). *For fixed k-graph $\mathcal{F}^{(k)}$ on f vertices, suppose $\mathcal{H}_n^{(k)}$ is a k-graph on n vertices containing $o(n^f)$ (not necessarily induced) copies of $\mathcal{F}^{(k)}$. Then, one may remove $o(n^k)$ edges from $\mathcal{H}_n^{(k)}$ to obtain a sub-hypergraph $\mathcal{G}^{(k)}$ which is $\mathcal{F}^{(k)}$-free, i.e., $\mathcal{G}^{(k)}$ contains no copy of $\mathcal{F}^{(k)}$ at all.*

When $\mathcal{F}^{(k)} = K_{k+1}^{(k)}$, the removal lemma generalizes Ruzsa and Szemerédi's triangle removal lemma (discussed earlier) to k-uniform hypergraphs. Frankl and the second author [FR02, Rödl91] observed that the the removal lemma (with $\mathcal{F}^{(k)} = K_{k+1}^{(k)}$) implies Szemerédi's theorem (see Theorem 1.4 below with $d = 1$). Subsequently, Solymosi [Sol03, Sol04] showed that Theorem 1.3 also implies the multidimensional version of Szemerédi's theorem, originally due to Furstenberg and Katznelson [FK78] (see also [RSTT06$^+$] for another consequence of Theorem 1.3 of similar flavor).

Theorem 1.4 (multidimensional Szemerédi theorem). *For fixed integers ℓ and d, any set $Z \subseteq \{1,\dots,n\}^d$ containing no homothetic copy of $\{1,\dots,\ell\}^d$ has size $|Z| = o(n^d)$.*

In Theorem 1.4, $\{1,\dots,n\}^d$ denotes, as usual, the d-fold cross product of the set $\{1,\dots,n\}$ with itself. A homothetic copy of $\{1,\dots,\ell\}^d$ is any set of the form $\boldsymbol{a} + c\{1,\dots,\ell\}^d$, where $\boldsymbol{a} \in \{1,\dots,n\}^d$ and c is some positive integer.

2 Results

In this paper, we consider some extremal hypergraph problems of Turán and Ramsey flavor. We begin with some problems of Turán-type.

2.1 A Turán-type Problem

Generalizing Turán's problem for hypergraphs, the following problem was initiated by Brown, Erdős and T. Sós [BES73]. Let $f^{(r)}(n, v, e)$ denote the maximum number of edges in an r-graph on n vertices in which no v vertices span e (or more) edges. Note that the determination of $f^{(r)}(n, v, \binom{v}{r}) = \mathrm{ex}(n, K_v^{(r)})$ is precisely Turán's problem, on which we shall expand in Section 2.2. It was first proved by Brown, Erdős and T. Sós [BES73] that $f^{(r)}(n, e(r-k)+k, e) = \Theta(n^k)$. The same authors asked what happens if, instead of on $v = e(r-k)+k$ vertices, one forbids e edges to appear on $v + 1 = e(r-k) + k + 1$ vertices. In particular, they conjectured $f^{(r)}(n, e(r-k)+k+1, e)$ can be bounded by $o(n^k)$. This conjecture was proved for $e = r = 3$ and $k = 2$ by Ruzsa and Szemerédi [RS78] and generalized to arbitrary r with $k = 2$ and $e = 3$ by Erdős, Frankl and Rödl [EFR86] and $r > k$ and $e = 3$ by Alon and Shapira [AS06$^+$]. Theorem 1.3 easily implies the upper bound for $r > k \geq 2$ and $e = k + 1$. We present the details in Section 3.

Theorem 2.1. *For $r > k \geq 2$, $f^{(r)}(n, (k+1)(r-k+1), k+1) = o(n^k)$.*

Theorem 2.1 was proved for $k = 2$ by Erdős, Frankl and Rödl [EFR86], and for $k = 3$ by Sárközy and Selkow [SS05].

Remark 2.2. In this paper, the integer notation k is usually reserved for the uniformity of hypergraphs $\mathcal{H}^{(k)}$, while our notation $f^{(r)}(n, v, e)$ appears to break with that tradition (since here, r denotes uniformity). In Theorem 2.1, however, the essential part of proving the assertion $f^{(r)}(n, (k+1)(r-k+1), k+1) = o(n^k)$, in fact, involves appealing to specific auxiliary *k-uniform* hypergraphs $\mathcal{H}^{(k)}$, where the initial uniformity r plays less of a rôle. In this sense, we reserve consistent use of uniformity notation k for later, in the proof, where we feel it is most important.

2.2 Forbidden Families

For an integer k, let $\mathbf{F}^{(k)} = \{\mathcal{F}_i^{(k)}\}_{i \in I}$ be a given (possibly infinite) family of k-graphs. Let $\mathrm{Forb}(n, \mathbf{F}^{(k)})$ denote the family of all k-graphs $\mathcal{H}_n^{(k)}$ on vertex set $\{1, \ldots, n\}$ containing no sub-hypergraph isomorphic to $\mathcal{F}_i^{(k)}$ for all $i \in I$. As in the classical Turán problem, set

$$\mathrm{ex}(n, \mathbf{F}^{(k)}) = \max\left\{ \left|\mathcal{H}_n^{(k)}\right| : \mathcal{H}_n^{(k)} \in \mathrm{Forb}(n, \mathbf{F}^{(k)}) \right\}.$$

When $\mathbf{F}^{(k)} = \{K_\ell^{(k)}\}$ consists of the single clique $K_\ell^{(k)}$, determining $\mathrm{ex}(n, K_\ell^{(k)}) = \mathrm{ex}(n, \{K_\ell^{(k)}\})$ is the well-known Turán problem, where even the asymptotic for the case $\ell = 4$ and $k = 3$ remains open today. For $k = 2$, Turán's formula for these numbers is a central result in extremal graph theory. Note that the parameter $\mathrm{ex}(n, K_\ell^{(k)})$ corresponds to $f^{(k)}(n, \ell, \binom{\ell}{k})$ from Section 2.1. In the context of Turán's problem, however, the 'ex' notation appears more commonly than the 'f' notation, and so we shall not break from this tradition here.

Our result in Theorem 2.3 below aims to relate $\mathrm{ex}(n, \mathbf{F}^{(k)})$ with the cardinality $|\mathrm{Forb}(n, \mathbf{F}^{(k)})|$. Observe that since all sub-hypergraphs of a fixed $\mathcal{H}_n^{(k)} \in \mathrm{Forb}(n, \mathbf{F}^{(k)})$ also belong to $\mathrm{Forb}(n, \mathbf{F}^{(k)})$, we have $|\mathrm{Forb}(n, \mathbf{F}^{(k)})| \geq 2^{\mathrm{ex}(n, \mathbf{F}^{(k)})}$. We show that this bound is, in a sense, best possible.

Theorem 2.3. *For every (possibly infinite) family of k-graphs $\mathbf{F}^{(k)} = \{\mathcal{F}_i^{(k)}\}_{i \in I}$, we have*

$$\log_2 |\mathrm{Forb}(n, \mathbf{F}^{(k)})| = \mathrm{ex}(n, \mathbf{F}^{(k)}) + o(n^k)\,.$$

Theorem 2.3 was proved for $k = 2$ and $\mathbf{F}^{(2)} = \{K_\ell^{(2)}\}$ by Erdős, Kleitman and Rothschild [EKR76] and for general $\mathbf{F}^{(2)}$ by Erdős, Frankl and Rödl [EFR86]. Theorem 2.3 was proved for $k = 3$ by the first two authors [NR98]. Bollobás and Thomason [BT95] showed that $\lim_{n\to\infty} \log_2 |\mathrm{Forb}(n, \mathbf{F}^{(k)})| / \binom{n}{k}$ exists for any family $\mathbf{F}^{(k)}$ and so Theorem 2.3 provides a combinatorial evaluation of this limit.

We mention that for $k = 2$, an induced version of Theorem 2.3 was established by Prömel and Steger [PS92] and by Bollobás and Thomason [BT97]. These results were extended to $k = 3$ by Kohayakawa and the first two authors in [KNR03]. Using the hypergraph regularity method, one may prove an induced version of Theorem 2.3 for general $k \geq 2$, and we hope to address this problem in a forthcoming paper.

2.3 An Induced Ramsey Theorem

For a fixed k-graph $\mathcal{F}^{(k)}$, a k-graph $\mathcal{G}^{(k)}$ is said to be an *induced Ramsey k-graph* for $\mathcal{F}^{(k)}$ if every 2-coloring of $\mathcal{G}^{(k)}$ admits a monochromatic sub-hypergraph isomorphic to $\mathcal{F}^{(k)}$ which appears as an *induced* sub-hypergraph of $\mathcal{G}^{(k)}$. Nešetřil and Rödl [NR77, NR82] and independently Abramson and Harrington [AH78] proved that every k-graph $\mathcal{F}^{(k)}$ has a Ramsey k-graph $\mathcal{G}^{(k)}$ for $\mathcal{F}^{(k)}$. In this paper, we present another proof of the induced Ramsey theorem (based on the hypergraph regularity method).

Theorem 2.4. *For every integer $k \geq 2$ and every k-graph $\mathcal{F}^{(k)}$, there exists an induced Ramsey k-graph $\mathcal{G}^{(k)}$ for $\mathcal{F}^{(k)}$.*

2.4 Organization of Paper

In Section 3, we prove Theorem 2.1 using the removal lemma, Theorem 1.3. While Theorem 2.1 is a consequence of the removal lemma, we prove Theorem 2.3 and Theorem 2.4 using the hypergraph regularity method. In Section 4, we present the hypergraph regularity lemma and hypergraph counting lemma. In Section 5, we prove Theorem 2.3. In Section 6, we prove Theorem 2.4.

We conclude the introduction with the following remark.

Remark 2.5. The components of the hypergraph regularity method, the hypergraph regularity lemma and hypergraph counting lemma, take different forms in the versions [Gow06$^+$, Gow06] and [NRS06, RS04] and subsequent versions [RS06$^+$] and [Tao06$^+$b]. While any of these versions would suffice to prove the applications in this paper, we find the recent version of this method due to the second two authors [RS06$^+$] (based on ideas from [NRS06, RS04]) most convenient for our purposes. We present these tools in Section 4.

3 Proof of Theorem 2.1

We use Theorem 1.3, the removal lemma, to prove Theorem 2.1. In particular, we use the following corollary of the removal lemma to prove Theorem 2.1.

Corollary 3.1. *For fixed integer $k \geq 2$, let k-graph $\mathcal{H}_n^{(k)}$ on n vertices have the property that each k-tuple $K \in \mathcal{H}_n^{(k)}$ belongs to precisely one copy of the clique $K_{k+1}^{(k)}$. Then, $|\mathcal{H}_n^{(k)}| = o(n^k)$.*

Proof. Corollary 3.1 follows easily from Theorem 1.3 in the case when $\mathcal{F}^{(k)}$ consists of the single k-clique $K_{k+1}^{(k)}$ on $k+1$ vertices.

Let $\mathcal{H}_n^{(k)}$ be given as in the hypothesis of Corollary 3.1. Since each k-tuple $K \in \mathcal{H}_n^{(k)}$ belongs to precisely one copy of $K_{k+1}^{(k)}$, we see that the number of such cliques, $|\mathcal{K}_{k+1}(\mathcal{H}_n^{(k)})|$, satisfies

$$\left|\mathcal{K}_{k+1}(\mathcal{H}_n^{(k)})\right| = \frac{1}{k+1}|\mathcal{H}_n^{(k)}| = o(n^{k+1}). \tag{1}$$

Putting $\mathcal{F}^{(k)} = K_{k+1}^{(k)}$, Theorem 1.3 then asserts that one may delete $o(n^k)$ many k-tuples $K \in \mathcal{H}_n^{(k)}$ to obtain a $K_{k+1}^{(k)}$-free sub-hypergraph $\widetilde{\mathcal{H}}_n^{(k)} \subseteq \mathcal{H}_n^{(k)}$. However, since deleting a k-tuple $K \in \mathcal{H}_n^{(k)}$ destroys exactly one clique $K_{k+1}^{(k)}$, we must have $|\mathcal{K}_{k+1}(\mathcal{H}_n^{(k)})| = o(n^k)$ and Corollary 3.1 follows from (1). □

Proof of Theorem 2.1. Our proof follows the lines of [EFR86, RS78, SS05], where the earlier established removal lemmas for graphs and 3-graphs were used to prove the special cases $k = 2, 3$. Let $r > k \geq 2$ be given as in Theorem 2.1. Suppose, on the contrary, that there exists $c = c(r,k) > 0$ and positive integer $n_0 = n_0(r,k,c)$ for which

$$f^{(r)}(n, (k+1)(r-k+1), k+1) > cn^k \tag{2}$$

holds for all $n > n_0$. Let $\mathcal{G}^{(r)}$ be an r-graph on $n > n_0(r,k,c)$ vertices with cn^k many r-tuples with the property that no $(k+1)(r-k+1)$ vertices span $(k+1)$

many r-tuples. We shall demonstrate that the existence of such $\mathcal{G}^{(r)}$ contradicts Corollary 3.1.

We begin by reducing the r-graph $\mathcal{G}^{(r)}$ to an r-partite sub-hypergraph $\widetilde{\mathcal{G}}^{(r)}$. A simple averaging argument (see, e.g., [EK68]) implies that the vertex set $V(\mathcal{G}^{(r)})$ admits an r-partition $V(\mathcal{G}^{(r)}) = V_1 \cup \ldots \cup V_r$ for which

$$\left|\mathcal{G}^{(r)}\left[V_1, \ldots, V_r\right]\right| \geq \frac{r!}{r^r}|\mathcal{G}^{(r)}| > \frac{r!}{r^r}cn^k, \tag{3}$$

where $\mathcal{G}^{(r)}[V_1, \ldots, V_r]$ is the sub-hypergraph of $\mathcal{G}^{(r)}$ consisting of all r-tuples $R \in \mathcal{G}^{(r)}$ with $|R \cap V_i| = 1$ for all $1 \leq i \leq r$. For simplicity, set $\widetilde{\mathcal{G}}^{(r)} = \mathcal{G}^{(r)}[V_1, \ldots, V_r]$.

We now reduce the r-graph $\widetilde{\mathcal{G}}^{(r)}$ to $(k+1)$-graph $\widetilde{\mathcal{G}}^{(k+1)}$ with vertex set $V_1 \cup \ldots \cup V_{k+1}$ as follows: for a $(k+1)$-tuple K^+ satisfying $|K^+ \cap V_i| = 1$, $1 \leq i \leq k+1$, put $K^+ \in \widetilde{\mathcal{G}}^{(k+1)}$ if, and only if, $K^+ \subseteq R$ for some $R \in \widetilde{\mathcal{G}}^{(r)}$. We make the following claim.

Claim 3.2. $|\widetilde{\mathcal{G}}^{(k+1)}| \geq \frac{|\widetilde{\mathcal{G}}^{(r)}|}{k} \overset{(3)}{>} \frac{cr!}{kr^r}n^k$.

Proof. The second inequality immediately follows from (3). To establish the first, we observe that for each $K^+ \in \widetilde{\mathcal{G}}^{(k+1)}$, there are at most k many r-tuples $R \in \widetilde{\mathcal{G}}^{(r)}$ for which $K^+ \subseteq R$ (from which Claim 3.2 then follows). Otherwise, if for some $K^+ \in \widetilde{\mathcal{G}}^{(k+1)}$, there exist $(k+1)$ distinct r-tuples $R_1, \ldots, R_{k+1} \in \widetilde{\mathcal{G}}^{(r)}$ each containing K^+, we would have $(k+1)$ many r-tuples spanned on

$$\left|\bigcup_{i=1}^{k+1} R_i\right| \leq (r-k-1)(k+1) + k + 1 = (r-k)(k+1) < (r-k+1)(k+1)$$

vertices, contradicting our choice of $\mathcal{G}^{(r)}$. □

We proceed with the following claim.

Claim 3.3. Let $K_0 \subset K_0^+ \in \widetilde{\mathcal{G}}^{(k+1)}$ with $|K_0| = k$. There are at most $k-1$ distinct $(k+1)$-tuples $K_1^+, \ldots, K_{k-1}^+ \in \widetilde{\mathcal{G}}^{(k+1)}$ for which $K_0^+ \cap K_i^+ = K_0$, $1 \leq i \leq k$.

Proof. Suppose, on the contrary, that some fixed $K_0 \subset K_0^+ \in \widetilde{\mathcal{G}}^{(k+1)}$, $|K_0| = k$, admits k distinct $(k+1)$-tuples $K_1^+, \ldots, K_k^+ \in \widetilde{\mathcal{G}}^{(k+1)}$ for which $K_0^+ \cap K_i^+ = K_0$, $1 \leq i \leq k$. Then, for some $R_0, R_1, \ldots, R_k \in \widetilde{\mathcal{G}}^{(r)}$, we would have $(k+1)$-many distinct r-tuples $R_0 \supset K_0^+, R_1 \supset K_1^+, \ldots, R_k \supset K_k^+$ spanned on

$$\left|\bigcup_{i=0}^{k+1} R_i\right| \leq k+k+1+(k+1)(r-k-1) = (k+1)(r-k+1)-1 < (k+1)(r-k+1)$$

vertices, contradicting our choice of $\mathcal{G}^{(r)}$. □

Claim 3.3 immediately implies that for each $K_0^+ \in \widetilde{\mathcal{G}}^{(k+1)}$, at most $(k+1)(k-1) = k^2-1$ distinct $(k+1)$-tuples $K_1^+, \ldots, K_{k^2-1}^+ \in \widetilde{\mathcal{G}}^{(k+1)}$ satisfy $|K_0^+ \cap K_i^+| = k$, $1 \le i \le k^2-1$. As such, the $(k+1)$-graph $\widetilde{\mathcal{G}}^{(k+1)}$ contains a sub-hypergraph $\widetilde{\mathcal{G}}_0^{(k+1)}$ of size

$$\left|\widetilde{\mathcal{G}}_0^{(k+1)}\right| \geq \frac{|\widetilde{\mathcal{G}}^{(k+1)}|}{k^2-1} \overset{\text{Claim 3.2}}{\geq} \frac{cr!}{k(k^2-1)r^r} n^k \tag{4}$$

consisting of $(k+1)$-tuples $K_0^+ \in \widetilde{\mathcal{G}}^{(k+1)}$, no two of which overlap in k vertices. Indeed, iteratively construct $(k+1)$-graph $\widetilde{\mathcal{G}}_0^{(k+1)}$ by starting with an arbitrary $(k+1)$-tuple $K_0^+ \in \widetilde{\mathcal{G}}^{(k+1)}$, deleting all $(k+1)$-tuples K^+ which overlap with K_0^+ in k vertices, and repeating this procedure until $\widetilde{\mathcal{G}}_0^{(k+1)}$ is produced.

We are now able to conclude the proof of Theorem 2.1. Define $(k+1)$-partite k-graph $\mathcal{H}^{(k)}$ on vertex set $V_1 \cup \ldots \cup V_{k+1}$ as follows: for a k-tuple K_0 satisfying $|K_0 \cap V_i| \le 1$, $1 \le i \le k+1$, put $K_0 \in \mathcal{H}^{(k)}$ if, and only if, $K_0 \subset K_0^+$ for some $(k+1)$-tuple $K_0^+ \in \widetilde{\mathcal{G}}_0^{(k+1)}$. We make the following observations.

(O1) each copy of the clique $K_{k+1}^{(k)}$ in $\mathcal{H}^{(k)}$ corresponds to an edge of $\widetilde{\mathcal{G}}_0^{(k+1)}$, and vice-versa;
(O2) by construction of $\mathcal{H}^{(k)}$, each edge $K \in \mathcal{H}^{(k)}$ belongs to at least one copy of the clique $K_{k+1}^{(k)}$ in $\mathcal{H}^{(k)}$;
(O3) by construction of $\widetilde{\mathcal{G}}_0^{(k+1)}$, each edge $K \in \mathcal{H}^{(k)}$ belongs to at most one copy of the clique $K_{k+1}^{(k)}$ in $\mathcal{H}^{(k)}$;
(O4)
$$|\mathcal{H}^{(k)}| = \binom{k+1}{k} \left|\widetilde{\mathcal{G}}_0^{(k+1)}\right| \overset{(4)}{\geq} \frac{cr!(k+1)}{k(k^2-1)r^r} n^k = \Omega(n^k).$$

Combining observations (O2), (O3) and (O4), we see that $\mathcal{H}^{(k)}$ is a 'dense' k-graph whose every edge $K \in \mathcal{H}^{(k)}$ belongs to precisely one copy of the clique $K_{k+1}^{(k)}$. This contradicts Corollary 3.1 and hence concludes the proof of Theorem 2.1. □

4 Regularity Method for Hypergraphs

In this section, we present the hypergraph regularity lemma and the hypergraph counting lemma from [RS06+]. We first present all needed definitions and notation in Section 4.1. In Section 4.2, we state both lemmas.

4.1 Definitions

We start with some basic concepts and notation.

Basic Concepts

For integers $\ell \geq j \geq 1$, the notation $[\ell]$ denotes the set of integers $\{1,\dots,\ell\}$ and $[\ell]^j = \binom{[\ell]}{j}$ denotes the set of all unordered j-tuples from $[\ell]$.

In this paper ℓ-partite, j-uniform hypergraphs play a special rôle, where $j \leq \ell$. Given vertex sets $V_1,\dots,V_\ell$, we denote by $K_\ell^{(j)}(V_1,\dots,V_\ell)$ the *complete* ℓ-partite, j-uniform hypergraph (i.e., the family of all j-element subsets $J \subseteq \bigcup_{i\in[\ell]} V_i$ satisfying $|V_i \cap J| \leq 1$ for every $i \in [\ell]$). If $|V_i| = m$ for every $i \in [\ell]$, then an (m,ℓ,j)*-cylinder* $\mathcal{H}^{(j)}$ on $V_1 \cup \ldots \cup V_\ell$ is any subset of $K_\ell^{(j)}(V_1,\dots,V_\ell)$. The vertex partition $V_1 \cup \ldots \cup V_\ell$ is an $(m,\ell,1)$-cylinder $\mathcal{H}^{(1)}$. (This definition may seem artificial right now, but it will simplify later notation.) For $j \leq i \leq \ell$ and set $\Lambda_i \in [\ell]^i$, we denote by $\mathcal{H}^{(j)}[\Lambda_i] = \mathcal{H}^{(j)}\big[\bigcup_{\lambda\in\Lambda_i} V_\lambda\big]$ the sub-hypergraph of the (m,ℓ,j)-cylinder $\mathcal{H}^{(j)}$ induced on $\bigcup_{\lambda\in\Lambda_i} V_\lambda$.

For an (m,ℓ,j)-cylinder $\mathcal{H}^{(j)}$ and an integer $2 \leq j \leq i \leq \ell$, we denote by $\mathcal{K}_i(\mathcal{H}^{(j)})$ the family of all i-element subsets of $V(\mathcal{H}^{(j)})$ which span complete sub-hypergraphs in $\mathcal{H}^{(j)}$ of order i. For $1 \leq i \leq \ell$, we denote by $\mathcal{K}_i(\mathcal{H}^{(1)})$ the family of all i-element subsets of $V(\mathcal{H}^{(1)})$ which 'cross' the partition $V_1 \cup \ldots \cup V_\ell$, i.e., $I \in \mathcal{K}_i(\mathcal{H}^{(1)})$ if, and only if, $|I \cap V_s| \leq 1$ for all $1 \leq s \leq \ell$. For $2 \leq j \leq i \leq \ell$, $|\mathcal{K}_i(\mathcal{H}^{(j)})|$ is the number of all copies of $K_i^{(j)}$ in $\mathcal{H}^{(j)}$. Given an $(m,\ell,j-1)$-cylinder $\mathcal{H}^{(j-1)}$ and an (m,ℓ,j)-cylinder $\mathcal{H}^{(j)}$, we say $\mathcal{H}^{(j-1)}$ *underlies* $\mathcal{H}^{(j)}$ if $\mathcal{H}^{(j)} \subseteq \mathcal{K}_j(\mathcal{H}^{(j-1)})$. This brings us to one of the main concepts of this paper, the notion of a *complex*.

Definition 4.1 ((m,ℓ,h)-complex). *Let $m \geq 1$ and $\ell \geq h \geq 1$ be integers. An (m,ℓ,h)-*complex *$\boldsymbol{\mathcal{H}}$ is a collection of (m,ℓ,j)-cylinders $\{\mathcal{H}^{(j)}\}_{j=1}^h$ such that*

(a) $\mathcal{H}^{(1)}$ is an $(m,\ell,1)$-cylinder, i.e., $\mathcal{H}^{(1)} = V_1 \cup \dots \cup V_\ell$ with $|V_i| = m$ for $i \in [\ell]$, and
(b) $\mathcal{H}^{(j-1)}$ underlies $\mathcal{H}^{(j)}$ for $2 \leq j \leq h$, i.e., $\mathcal{H}^{(j)} \subseteq \mathcal{K}_j(\mathcal{H}^{(j-1)})$.

We sometimes shorten the terminology (m,ℓ,h)-complex to (ℓ,h)*-complex*, when the cardinality $m = |V_1| = \dots = |V_s|$ isn't of primary concern.

Relative Density and Hypergraph Regularity

We begin by defining a relative density of a j-uniform hypergraph w.r.t. $(j-1)$-uniform hypergraph on the same vertex set.

Definition 4.2 (relative density). *Let $\mathcal{H}^{(j)}$ be a j-uniform hypergraph and let $\mathcal{H}^{(j-1)}$ be a $(j-1)$-uniform hypergraph on the same vertex set. We define the* density *of $\mathcal{H}^{(j)}$* w.r.t. *$\mathcal{H}^{(j-1)}$ as*

$$d\big(\mathcal{H}^{(j)}\big|\mathcal{H}^{(j-1)}\big) = \begin{cases} \frac{\left|\mathcal{H}^{(j)}\cap\mathcal{K}_j(\mathcal{H}^{(j-1)})\right|}{\left|\mathcal{K}_j(\mathcal{H}^{(j-1)})\right|} & \text{if} \quad \left|\mathcal{K}_j(\mathcal{H}^{(j-1)})\right| > 0 \\ 0 & \text{otherwise}. \end{cases}$$

We also define a notion of regularity for (m, j, j)-cylinders w.r.t. some underlying $(m, j, j-1)$-cylinders.

Definition 4.3 ((ε, d)-regular). *Let reals $\varepsilon > 0$ and $d \geq 0$ be given along with an (m, j, j)-cylinder $\mathcal{H}^{(j)}$ and underlying $(m, j, j-1)$-cylinder $\mathcal{H}^{(j-1)}$. We say $\mathcal{H}^{(j)}$ is (ε, d)-*regular *w.r.t. $\mathcal{H}^{(j-1)}$ if whenever $\mathcal{Q}^{(j-1)} \subseteq \mathcal{H}^{(j-1)}$ satisfies*

$$\big|\mathcal{K}_j(\mathcal{Q}^{(j-1)})\big| \geq \varepsilon \big|\mathcal{K}_j(\mathcal{H}^{(j-1)})\big|, \quad \textit{then} \quad d\big(\mathcal{H}^{(j)} \big| \mathcal{Q}^{(j-1)}\big) = d \pm \varepsilon\,.$$

Before continuing, we pause for the following remark.

Remark 4.4. We compare the notion of regularity in Definition 4.3 for $j = 2$ with the traditional definition of an ε-regular pair (given in the beginning of the Introduction). The $(m, 2, 2)$-cylinder $\mathcal{H}^{(2)}$ is, in the traditional terminology, a bipartite graph. The underlying $(m, 2, 1)$-cylinder $\mathcal{H}^{(1)}$ is the bipartition of $\mathcal{H}^{(2)}$, written here as $\mathcal{H}^{(1)} = V_1 \cup V_2$ where $|V_1| = |V_2| = m$. The sub-cylinder $\mathcal{Q}^{(1)} \subseteq V_1 \cup V_2$ is a subset of vertices, which we could write as $\mathcal{Q}^{(1)} = V_1' \cup V_2'$, where $V_1' \subseteq V_1$ and $V_2' \subseteq V_2$. The assumption of Definition 4.3 saying $|\mathcal{K}_2(\mathcal{Q}^{(1)})| \geq \varepsilon |\mathcal{K}_2(\mathcal{H}^{(1)})|$ is identical to saying $|V_1'||V_2'| \geq \varepsilon |V_1||V_2|$. As such, the definition ensures $d(\mathcal{H}^{(2)}|\mathcal{Q}^{(1)}) = d \pm \varepsilon$, or equivalently, $|d(\mathcal{H}^{(2)}|\mathcal{Q}^{(1)}) - d| < \varepsilon$. The quantity $d(\mathcal{H}^{(2)}|\mathcal{Q}^{(1)})$ is the same as $d_{\mathcal{H}^{(2)}}(V_1', V_2')$. The constant d is not necessarily the density $d(\mathcal{H}^{(2)}|\mathcal{H}^{(1)})$, but it is, of course, close to it.

There is only one real difference, therefore, between the notion of graph regularity given in Definition 4.3 when $j = 2$ and the traditional definition of an ε-regular pair. In the traditional definition, we would assume that the subsets $V_1' \subseteq V_1$, $V_2' \subseteq V_2$ individually satisfy the conditions $|V_1'| \geq \varepsilon |V_1|$ and $|V_2'| \geq \varepsilon |V_2|$. In Definition 4.3, we assume the product $|V_1'||V_2'|$ satisfies the single condition $|V_1'||V_2'| \geq \varepsilon |V_1||V_2|$. Quite obviously, however, these two notions are equivalent: *if $\mathcal{H}^{(2)}$ is (ε, d)-regular w.r.t. $\mathcal{H}^{(1)}$, then $\mathcal{H}^{(1)}$ is an ε-regular pair, and if $\mathcal{H}^{(1)}$ is an ε-regular pair, then $\mathcal{H}^{(2)}$ is $(\varepsilon^2, d(\mathcal{H}^{(2)}, \mathcal{H}^{(1)}))$-regular w.r.t. $\mathcal{H}^{(1)}$.*

We now extend the notion of (ε, d)-regularity to (m, ℓ, j)-cylinders $\mathcal{H}^{(j)}$.

Definition 4.5 ((ε, d)-regular cylinder). *We say an (m, ℓ, j)-cylinder $\mathcal{H}^{(j)}$ is (ε, d)-*regular *w.r.t. an $(m, \ell, j-1)$-cylinder $\mathcal{H}^{(j-1)}$ if for every $\Lambda_j \in [\ell]^j$, the restriction $\mathcal{H}^{(j)}[\Lambda_j] = \mathcal{H}^{(j)}\big[\bigcup_{\lambda \in \Lambda_j} V_\lambda\big]$ is (ε, d)-regular w.r.t. the restriction $\mathcal{H}^{(j-1)}[\Lambda_j] = \mathcal{H}^{(j-1)}\big[\bigcup_{\lambda \in \Lambda_j} V_\lambda\big]$.*

We now extend the notion of (ε, d)-regularity from cylinders to complexes.

Definition 4.6 ($(\varepsilon, \boldsymbol{d})$-regular complex). *Let ε be a positive real and let $\boldsymbol{d} = (d_2, \ldots, d_h)$ be a vector of non-negative reals. We say an (m, ℓ, h)-complex $\boldsymbol{\mathcal{H}} = \{\mathcal{H}^{(j)}\}_{j=1}^h$ is $(\varepsilon, \boldsymbol{d})$-*regular *if $\mathcal{H}^{(j)}$ is (ε, d_j)-regular w.r.t. $\mathcal{H}^{(j-1)}$ for every $j = 2, \ldots, h$.*

Partitions

The regularity lemma for k-uniform hypergraphs provides a well-structured family of partitions $\mathscr{P} = \{\mathscr{P}^{(1)}, \ldots, \mathscr{P}^{(k-1)}\}$ of vertices, pairs, ..., and $(k-1)$-tuples of some vertex set. We now discuss the structure of these partitions recursively, following the approach of [RS04].

Let k be a fixed integer and V be a set of vertices. Let $\mathscr{P}^{(1)} = \{V_1, \ldots, V_{|\mathscr{P}^{(1)}|}\}$ be a partition of V. For every $1 \leq j \leq |\mathscr{P}^{(1)}|$, let $\mathrm{Cross}_j(\mathscr{P}^{(1)})$ be the family of all crossing j-tuples J, i.e., the set of j-tuples which satisfy $|J \cap V_i| \leq 1$ for every $V_i \in \mathscr{P}^{(1)}$.

Suppose that partitions $\mathscr{P}^{(i)}$ of $\mathrm{Cross}_i(\mathscr{P}^{(1)})$ for $1 \leq i \leq j-1$ have been defined. Then for every $(j-1)$-tuple I in $\mathrm{Cross}_{j-1}(\mathscr{P}^{(1)})$, there exist a unique $\mathcal{P}^{(j-1)} = \mathcal{P}^{(j-1)}(I) \in \mathscr{P}^{(j-1)}$ so that $I \in \mathcal{P}^{(j-1)}$. For every j-tuple J in $\mathrm{Cross}_j(\mathscr{P}^{(1)})$, we define the *polyad of J*

$$\hat{\mathcal{P}}^{(j-1)}(J) = \bigcup \left\{ \mathcal{P}^{(j-1)}(I) \colon I \in [J]^{j-1} \right\} .$$

In other words, $\hat{\mathcal{P}}^{(j-1)}(J)$ is the unique set of j partition classes of $\mathscr{P}^{(j-1)}$ each containing a $(j-1)$-subset of J. Observe that $\hat{\mathcal{P}}^{(j-1)}(J)$ can be viewed as a $(j, j-1)$-cylinder, i.e., a j-partite, $(j-1)$-uniform hypergraph. More generally, for $1 \leq i < j$, we set

$$\hat{\mathcal{P}}^{(i)}(J) = \bigcup \left\{ \mathcal{P}^{(i)}(I) \colon I \in [J]^{i} \right\} \qquad \text{and} \qquad \boldsymbol{\mathcal{P}}(J) = \left\{ \hat{\mathcal{P}}^{(i)}(J) \right\}_{i=1}^{j-1} . \tag{5}$$

Remark 4.7. In this paper, we use $\mathscr{P}^{(j)}$, read “script P”, to denote the partition of j-tuples. Partition classes $\mathcal{P}^{(j)} \in \mathscr{P}^{(j)}$ (which are j-uniform hypergraphs on $[n]$) are denoted with “calligraphic P”. Unions of special sub-collections of j-graphs $\mathcal{P}^{(j)}$ (which we call polyads) are denoted with “calligraphic P” equipped with a “hat”.

Next, we define $\hat{\mathscr{P}}^{(j-1)}$, the family of all polyads

$$\hat{\mathscr{P}}^{(j-1)} = \left\{ \hat{\mathcal{P}}^{(j-1)}(J) \colon J \in \mathrm{Cross}_j(\mathscr{P}^{(1)}) \right\} .$$

Note that $\hat{\mathcal{P}}^{(j-1)}(J)$ and $\hat{\mathcal{P}}^{(j-1)}(J')$ are not necessarily distinct for different j-tuples J and J'. We view $\hat{\mathscr{P}}^{(j-1)}$ as a set and, consequently, $\{\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)}) \colon \hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}\}$ is a partition of $\mathrm{Cross}_j(\mathscr{P}^{(1)})$.

The structural requirement on the partition $\mathscr{P}^{(j)}$ of $\mathrm{Cross}_j(\mathscr{P}^{(1)})$ is

$$\mathscr{P}^{(j)} \prec \{\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)}) \colon \hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}\} , \tag{6}$$

where ‘$\prec$’ denotes the refinement relation of set partitions. In other words, we require that the set of cliques spanned by a polyad in $\hat{\mathscr{P}}^{(j-1)}$ is subpartitioned in $\mathscr{P}^{(j)}$ and every partition class in $\mathscr{P}^{(j)}$ belongs to precisely one polyad in $\hat{\mathscr{P}}^{(j-1)}$. Note that (6) implies (inductively) that $\boldsymbol{\mathcal{P}}(J)$ defined in (5) is a $(j, j-1)$-complex. On a related note, we shall often drop the argument $J \in \mathrm{Cross}_j(\mathscr{P}^{(1)})$ from the notation $\hat{\mathcal{P}}^{(j-1)}(J)$ (as the families $\mathscr{P}$ with which we work always satisfy $\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)}) \neq \emptyset$).

Throughout this paper, we want to control the number of partition classes in $\mathscr{P}^{(j)}$, and more specifically, over the number of classes contained in $\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})$ for a fixed polyad $\hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}$. We make this precise in the following definition.

Definition 4.8 (family of partitions). *Suppose V is a set of vertices, $k \geq 2$ is an integer and $\boldsymbol{a} = (a_1, \dots, a_{k-1})$ is a vector of positive integers. We say $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a}) = \{\mathscr{P}^{(1)}, \dots, \mathscr{P}^{(k-1)}\}$ is a* family of partitions on V, *if it satisfies the following:*

(i) $\mathscr{P}^{(1)}$ *is a partition of V into a_1 classes,*
(ii) $\mathscr{P}^{(j)}$ *is a partition of* $\mathrm{Cross}_j(\mathscr{P}^{(1)})$ *satisfying:*

$$\mathscr{P}^{(j)} \quad \textit{refines} \quad \{\mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)}) \colon \hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}\}$$
$$\textit{and} \quad \left|\left\{\mathcal{P}^{(j)} \in \mathscr{P}^{(j)} \colon \mathcal{P}^{(j)} \subseteq \mathcal{K}_j(\hat{\mathcal{P}}^{(j-1)})\right\}\right| = a_j \quad \textit{for every} \quad \hat{\mathcal{P}}^{(j-1)} \in \hat{\mathscr{P}}^{(j-1)}.$$

Moreover, we say $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a})$ is t-bounded, if $\max\{a_1, \dots, a_{k-1}\} \leq t$.

It is easy to see that for a t-bounded family of partitions $\mathscr{P}$ and an integer $2 \leq j \leq k-1$, we have

$$|\hat{\mathscr{P}}^{(j-1)}| = \binom{a_1}{j} \prod_{h=2}^{j-1} a_h^{\binom{j}{h}} \leq t^{2^t}. \tag{7}$$

We continue with a few final definitions needed to state the hypergraph regularity lemma and corresponding counting lemma.

Regular Partitions

The following definition describes some of the structure the regularity lemma shall provide.

Definition 4.9 ($(\eta, \varepsilon, \boldsymbol{a})$-equitable). *Suppose V is a set of n vertices, η and ε are positive reals, $\boldsymbol{a} = (a_1, \dots, a_{k-1})$ is a vector of positive integers where a_1 divides n.*

We say a family of partitions $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a})$ on V (as defined in Definition 4.8) is $(\eta, \varepsilon, \boldsymbol{a})$-equitable *if it satisfies the following:*

(a) $\left|[V]^k \setminus \mathrm{Cross}_k(\mathscr{P}^{(1)})\right| \leq \eta\binom{n}{k}$,
(b) $\mathscr{P}^{(1)} = \{V_i \colon i \in [a_1]\}$ *is an equitable vertex partition, i.e.,* $|V_i| = |V|/a_1$ *for* $i \in [a_1]$, *and*
(c) for every $K \in \mathrm{Cross}_k(\mathscr{P}^{(1)})$ *the* $(n/a_1, k, k-1)$*-complex* $\boldsymbol{\mathcal{P}}(K)$ *(see (5)) is* $(\varepsilon, (1/a_2, \dots, 1/a_{k-1}))$*-regular.*

To describe the remaining structure of the regularity lemma, we extend Definition 4.5.

Definition 4.10 ((δ_k, d_k, r)-regular). *Let δ_k and d_k be positive reals and r be a positive integer. Suppose $\mathcal{H}^{(k-1)}$ is a $(k-1)$-graph and $\mathcal{H}^{(k)}$ is a k-graph, both of which share the same vertex set. We say $\mathcal{H}^{(k)}$ is (δ_k, d_k, r)-*regular *w.r.t. $\mathcal{H}^{(k-1)}$ if for every collection $\boldsymbol{\mathcal{Q}}^{(k-1)} = \{\mathcal{Q}_1^{(k-1)}, \ldots, \mathcal{Q}_r^{(k-1)}\}$ of not necessarily disjoint sub-hypergraphs of $\mathcal{H}^{(k-1)}$ satisfying*

$$\left|\bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})\right| > \delta_k \left|\mathcal{K}_k(\mathcal{H}^{(k-1)})\right| ,$$

we have

$$\frac{\left|\mathcal{H}^{(k)} \cap \bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})\right|}{\left|\bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})\right|} = d_k \pm \delta_k .$$

We write $(\delta_k, *, r)$-regular *to mean $\big(\delta_k, d\big(\mathcal{H}^{(k)}\big|\mathcal{H}^{(k-1)}\big), r\big)$-regular.*

We need one last definition to state the regularity lemma.

Definition 4.11 ((δ_k, r)-regular w.r.t. $\mathscr{P}$). *Suppose δ_k is a positive real and r is a positive integer. Let $\mathcal{H}^{(k)}$ be a k-uniform hypergraph with vertex set V and $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a})$ be a family of partitions on V. We say $\mathcal{H}^{(k)}$* is (δ_k, r)-regular w.r.t. $\mathscr{P}$, *if*

$$\left|\bigcup \Big\{\mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}) \colon \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)} \text{ and } \mathcal{H}^{(k)} \text{ is not } (\delta_k, *, r)\text{-regular w.r.t. } \hat{\mathcal{P}}^{(k-1)}\Big\}\right| \le \delta_k \binom{|V|}{k} .$$

4.2 Hypergraph Regularity Lemma and Counting Lemma

The regularity lemma of [RS06$^+$] is given as follows.

Theorem 4.12 (Regularity lemma). *Let $k \ge 2$ be a fixed integer. For all positive constants η and δ_k and functions $r\colon \mathbb{N}^{k-1} \to \mathbb{N}$ and $\delta\colon \mathbb{N}^{k-1} \to (0,1]$ there are integers $t_{\text{Thm.4.12}}$ and $n_{\text{Thm.4.12}}$ so that the following holds.*

For every k-uniform hypergraph $\mathcal{H}^{(k)}$ satisfying $|V(\mathcal{H}^{(k)})| = n \ge n_{\text{Thm.4.12}}$ and $t_{\text{Thm.4.12}}!$ dividing n, there exists a family of partitions $\mathscr{P} = \mathscr{P}(k-1, \boldsymbol{a}^{\mathscr{P}})$ so that

(i) $\mathscr{P}$ is $(\eta, \delta(\boldsymbol{a}^{\mathscr{P}}), \boldsymbol{a}^{\mathscr{P}})$-equitable and $t_{\text{Thm.4.12}}$-bounded;
(ii) $\mathcal{H}^{(k)}$ is $(\delta_k, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\mathscr{P}$.

The following hypergraph counting lemma corresponds to Theorem 4.12.

Theorem 4.13 (Counting lemma). *For all integers $\ell \geq k \geq 2$ and positive constants $\gamma > 0$ and $d_k > 0$, there exists $\delta_k > 0$ such that for all integers $a_{k-1}, \dots, a_2$, there are a constant $\delta > 0$ and positive integers r and m_0 so that the following holds. Suppose*

(i) $\boldsymbol{\mathcal{R}} = \{\mathcal{R}^{(j)}\}_{j=1}^{k-1}$ is a $(\delta, (1/a_2, \dots, 1/a_{k-1}))$-regular $(m, \ell, k-1)$-complex with $m \geq m_0$, and
(ii) for every $\Lambda_k \in [\ell]^k$, the k-graph $\mathcal{H}^{(k)} \subseteq \mathcal{K}_k(\mathcal{R}^{(k-1)})$ is $(\delta_k, d_{\Lambda_k}, r)$-regular w.r.t. $\mathcal{R}^{(k-1)}[\Lambda_k]$ for some $d_{\Lambda_k} \geq d_k$.

Then

$$\left|\mathcal{K}_\ell(\mathcal{H}^{(k)})\right| \geq (1-\gamma) d_k^{\binom{\ell}{k}} \prod_{j=2}^{k-1} \left(\frac{1}{a_j}\right)^{\binom{\ell}{j}} \times m^\ell .$$

5 Proof of Theorem 2.3

The main idea in proving Theorem 2.3 is not difficult, but since it involves appealing to the regularity lemma and counting lemma for hypergraphs, its appearance is technical. We therefore begin this section by sketching this main idea in the (more transparent) case of graphs, following the work of [EFR86]. In the following outline, we restrict our attention to the special case when $\mathbf{F}^{(2)} = \{K_3^{(2)}\}$ consists of the (single) triangle $K_3 = K_3^{(2)}$. We mention that, if we focus our attention to when $\mathbf{F}^{(2)}$ consists of a single graph, our choice here of K_3 makes little difference in the argument. However, restricting our attention to when $\mathbf{F}^{(2)}$ consists of only *finitely* many graphs frees us from one detail which is similarly technical for graphs as it is for hypergraphs.

5.1 The Graph Case with $\mathbf{F}^{(2)} = K_3$

Fix $\nu > 0$. We sketch the proof that

$$\left|\mathrm{Forb}(n, K_3)\right| \leq 2^{\mathrm{ex}(n,K_3) + \nu n^3}$$

holds for all large integers n. The main components of the proof are the Szemerédi regularity lemma, Theorem 1.1, and the counting lemma (for graphs), Fact 1.2.

We begin by 'regularizing' every graph $G = \mathcal{G}^{(2)}$ in the collection $\mathrm{Forb}(n, K_3)$. To that end, we pick some 'small' $0 < \varepsilon = \varepsilon(\nu) \ll \nu$ (we won't determine a formula for ε at this time since we plan to bypass, in this outline, the calculations using this formula) and 'large' integer $t_0 = t_0(\nu) \gg 1/\nu$. As we make these choices, we also pick an auxiliary constant $\varepsilon \ll d_0 \ll \nu$ which is 'small' w.r.t. ν but 'large' w.r.t. ε. Theorem 1.1 guarantees an integer $T_0 = T_0(\varepsilon, t_0)$ so that, with n large, every graph $G \in \mathrm{Forb}(n, K_3)$ admits an ε-regular, t_G-equitable partition $V(G) = V_1^G \cup \ldots \cup V_{t_G}^G$ where $t_0 \leq t_G \leq T_0$. For $G \in \mathrm{Forb}(n, K_3)$, we shall write $\mathcal{P}_G$ for the ε-regular, t_G-equitable partition $V(G) = V_1^G \cup \ldots \cup V_{t_G}^G$, $t_0 \leq t_G \leq T_0$, obtained above. We fix, for each $G \in \mathrm{Forb}(n, K_3)$, the partition $\mathcal{P}_G$ now obtained (and if G admits multiple such, we simply pick one, arbitrarily). In all that follows, $n = n(\nu, d_0, \varepsilon, t_0, T_0)$ is sufficiently large w.r.t. all the constants mentioned above.

We first decompose $\mathrm{Forb}(n, K_3)$ into equivalence classes. We say two graphs G_1 and $G_2 \in \mathrm{Forb}(n, K_3)$ are equivalent if, and only if, $\mathcal{P}_{G_1} = \mathcal{P}_{G_2}$. (In other words, the ε-regular partitions $\mathcal{P}_{G_1}$ and $\mathcal{P}_{G_2}$ fixed above split the vertices $\{1, \ldots, n\}$ in precisely the same way.) Let $\mathrm{Forb}(n, K_3) = \Pi_1 \cup \ldots \cup \Pi_N$ be the partition of $\mathrm{Forb}(n, K_3)$ associated with this equivalence relation. Then $\left|\mathrm{Forb}(n, K_3)\right| = \sum_{a=1}^{N} |\Pi_a|$, and clearly, there are at most $N \leq T_0^n = 2^{o(n^2)}$ partitions of the vertices $\{1, \ldots, n\}$. Thus, it suffices to estimate $|\Pi_a|$ for an arbitrary index $1 \leq a \leq N$.

For the remainder of this outline, fix $1 \leq a \leq N$. There is a common partition $\mathcal{P}_a$ of $\{1, \ldots, n\}$ that every graph $G \in \Pi_a$ admits as its fixed ε-regular partition $\mathcal{P}_G$. We write $\mathcal{P}_a$ as $V_1 \cup \ldots \cup V_t$, where $t_0 \leq t \leq T_0$. Now, for a fixed $G \in \Pi_a$, we shall record for which pairs (V_i, V_j) of the partition $\mathcal{P}_a$ the graph G is 'dense' and 'regular'. More formally, for $1 \leq i < j \leq t$, write $\boldsymbol{x}_G = (x_{ij}^G : 1 \leq i < j \leq t)$, where

$$x_{ij}^G = \begin{cases} 1 \text{ if } \quad d_G(V_i, V_j) \geq d_0 \text{ and } V_i, V_j \text{ is } \varepsilon\text{-regular w.r.t. } G, \\ 0 \text{ otherwise.} \end{cases}$$

For fixed $\boldsymbol{x} \in \{0,1\}^{\binom{t}{2}}$, we set $\Pi_a(\boldsymbol{x}) = \{G \in \Pi_a : \boldsymbol{x}_G = \boldsymbol{x}\}$ and observe

$$|\Pi_a| = \sum \left\{ |\Pi_a(\boldsymbol{x})| \colon \boldsymbol{x} \in \{0,1\}^{\binom{t}{2}} \right\}.$$

Since there are only $2^{\binom{t}{2}} \leq 2^{T_0^2} = 2^{O(1)} = 2^{o(n^2)}$ vectors $\boldsymbol{x} \in \{0,1\}^{\binom{t}{2}}$, it suffices to estimate $|\Pi_a(\boldsymbol{x})|$ for a fixed but arbitrary $\boldsymbol{x} \in \{0,1\}^{\binom{t}{2}}$.

With $\boldsymbol{x}$ fixed, and a fixed before, we now define $D_a(\boldsymbol{x})$ as the graph with vertex set $\{1, \ldots, t\}$ and edges $\{i, j\}$, $1 \leq i < j \leq t$, corresponding to when the pair (V_i, V_j) is 'dense' and 'regular' w.r.t. *every* graph $G \in \Pi_a(\boldsymbol{x})$, i.e., when $x_{ij} = 1$. If we can show

$$|D_a(\boldsymbol{x})| \leq \mathrm{ex}(t, K_3) \tag{8}$$

then it will be easy to show

$$|\Pi_a(\boldsymbol{x})| \leq 2^{\mathrm{ex}(n, K_3) + \frac{\nu}{2} n^2}. \tag{9}$$

Establishing the implication (8) $\Longrightarrow$ (9) is standard, and so we only highlight it here. Indeed, using standard considerations of ε-regular partitions, one may easily show that for any $G \in \Pi_a(\boldsymbol{x})$

$$\Big|\big\{\{v_i,v_j\}\in E(G)\colon v_i\in V_i,\, v_j\in V_j,\ \text{ either } i=j \text{ or } x_{ij}=0\big\}\Big| \\ < \left(\frac{1}{t_0}+\varepsilon+d_0\right)n^2 \ll \frac{\nu}{2}n^2 \tag{10}$$

where the last 'inequality' holds by virtue of the fact that we chose $1/t_0$, ε, and d_0 much smaller than ν. Hence there are essentially $2^{\frac{\nu}{2}n^2}$ choices for the subgraphs of graphs $G\in\Pi_a(\boldsymbol{x})$ induced on vertex classes V_i $(i=1,\dots,t)$ and on pairs (V_i,V_j) with $x_{ij}=0$. The number of subgraphs on pairs (V_i,V_j) with $x_{ij}=1$ is (ignoring precise error calculations) approximately

$$2^{\sum_{\{i,j\}\in D_a(\boldsymbol{x})}|V_i||V_j|} \sim 2^{\sum_{\{i,j\}\in D_a(\boldsymbol{x})}\frac{n^2}{t^2}} \overset{(8)}{\lesssim} 2^{\mathrm{ex}(n,K_3)} \tag{11}$$

where the last asymptotic employs (8) and makes use of the fact that $\mathrm{ex}(t,K_3)/\binom{t}{2} \sim \mathrm{ex}(n,K_3)/\binom{n}{2}$ whenever t and n are large (recall $t\ge t_0$, where we picked t_0 'large'). Since every graph $G\in\Pi_a(\boldsymbol{x})$ behaves 'identically' on the common partition $\mathcal{P}_a$, every graph $G\in\Pi_a(\boldsymbol{x})$ must consist of one of the (essentially) $2^{\frac{\nu}{2}n^2}$ many subgraphs counted in (10), and one of the (essentially) $2^{\mathrm{ex}(n,K_3)}$ subgraphs counted in (11). This completes the sketch of (8) $\Longrightarrow$ (9).

We finish the present outline by proving (8), and to that end, we use the counting lemma, Fact 1.2. Indeed, if $|D_a(\boldsymbol{x})|>\mathrm{ex}(t,K_3)$, then $D_a(\boldsymbol{x})$ contains a copy of the triangle K_3. Let i,j,k denote the vertices of this triangle (which correspond to the vertex classes V_i,V_j,V_k of the partition $\mathcal{P}_a$) and fix any graph $G_0\in\Pi_a(\boldsymbol{x})$. By definition of $D_a(\boldsymbol{x})$, each of the pairs $\{V_i,V_j\}$, $\{V_j,V_k\}$ and $\{V_i,V_k\}$ are ε-regular w.r.t. G_0 and also satisfy

$$d_{G_0}(V_i,V_j),\, d_{G_0}(V_j,V_k),\, d_{G_0}(V_i,V_k)\ge d_0.$$

By the counting lemma, Fact 1.2, the graph G_0 contains at least $\sim d_0^3(n/t)^3>0$ many triangles K_3, which contradicts that $G_0\in\mathrm{Forb}(n,K_3)$. This completes the outline.

Before proceeding to the actual proof of Theorem 2.3, we make the following remark.

Remark 5.1. As we mentioned before, one has to work a little harder, whether for graphs or hypergraphs, when the set $\mathbf{F}^{(k)}$ consists of infinitely many elements rather than finitely many. These details were not addressed in our outline, but are addressed in our proof of Theorem 2.3. As well, in our proof of Theorem 2.3, we shall define a k-graph $\mathcal{D}_\alpha(\boldsymbol{x})$ in (26) which is an analogue to the graph $D_a(\boldsymbol{x})$ (cf. (8)). For reasons we do not mention here, we define $\mathcal{D}_\alpha(\boldsymbol{x})$ in a slightly different way than we defined $D_a(\boldsymbol{x})$. In the end, however, the invocation of the counting lemma will be precisely the same as in the outline above.

5.2 Setting Up the Proof of Theorem 2.3

In our proof of Theorem 2.3, we use the following notation. For an integer n and a family of k-graphs $\mathbf{F}^{(k)}$, set

$$\widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}) = \frac{\mathrm{ex}(n, \mathbf{F}^{(k)})}{\binom{n}{k}}.$$

It is well known (see [KNS64]) that the sequence $(\widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}))_{n=1}^{\infty}$ is non-increasing, and hence,

$$\pi(\mathbf{F}^{(k)}) = \lim_{n\to\infty} \widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}) \tag{12}$$

exists. Note that when $\pi(\mathbf{F}^{(k)}) = 0$ the assertion of Theorem 2.3 is trivial. Indeed,

$$|\mathrm{Forb}(n, \mathbf{F}^{(k)})| \le \sum_{s=0}^{o(n^k)} \binom{\binom{n}{k}}{s} = 2^{o(n^k)}.$$

Henceforth, we shall assume $\pi(\mathbf{F}^{(k)}) > 0$.

It suffices to prove Theorem 2.3 for n divisible by a fixed but arbitrary integer T. In particular, suppose that, for fixed $\nu > 0$ and fixed integer T, for every integer $m > m_0(k, \nu, T)$, we have

$$|\mathrm{Forb}(mT, \mathbf{F}^{(k)})| \le 2^{\mathrm{ex}(mT, \mathbf{F}^{(k)}) + \nu(mT)^k}.$$

Then it easily follows that for all integers $n > n_0(k, \nu, T)$,

$$|\mathrm{Forb}(n, \mathbf{F}^{(k)})| \le 2^{\mathrm{ex}(n, \mathbf{F}^{(k)}) + 2\nu n^k}.$$

Indeed, for an integer n, write $(m-1)T \le n < mT$ for some integer m. Then, with m and n sufficiently large, we have

$$\begin{aligned}
\log_2 |\mathrm{Forb}(n, \mathbf{F}^{(k)})| &\le \log_2 |\mathrm{Forb}(mT, \mathbf{F}^{(k)})| \le \mathrm{ex}(mT, \mathbf{F}^{(k)}) + \nu(mT)^k \\
&= \widetilde{\mathrm{ex}}(mT, \mathbf{F}^{(k)}) \binom{mT}{k} + \nu(mT)^k \le \pi(\mathbf{F}^{(k)}) \binom{mT}{k} + \nu(mT)^k + o((mT)^k) \\
&\le \pi(\mathbf{F}^{(k)}) \binom{n+T}{k} + \nu(n+T)^k + o((n+T)^k) = \pi(\mathbf{F}^{(k)}) \binom{n}{k} + \nu n^k + o(n^k) \\
&\le \widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}) \binom{n}{k} + \nu n^k + o(n^k) \le \widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}) \binom{n}{k} + 2\nu n^k \\
&= \mathrm{ex}(n, \mathbf{F}^{(k)}) + 2\nu n^k,
\end{aligned}$$

where the next to last inequality follows from the sequence $(\widetilde{\mathrm{ex}}(s, \mathbf{F}^{(k)}))_{s=1}^{\infty})$ being non-increasing with limit $\pi(\mathbf{F}^{(k)})$.

We now prove that for every $\nu > 0$, there exist integers $T = T(\nu)$ and $n_0 = n_0(\nu, T)$ so that for every $n \ge n_0$ divisible by T,

$$\log_2 |\mathrm{Forb}(n, \mathbf{F}^{(k)})| \le \mathrm{ex}(n, \mathbf{F}^{(k)}) + \nu \binom{n}{k} . \qquad (13)$$

As our proof depends on Theorems 4.12 and 4.13, we first discuss a sequence of auxiliary constants.

5.3 Constants

Let $\nu > 0$ be given. Let $f_0 \in \mathbb{N}$ be sufficiently large so that

$$\widetilde{\mathrm{ex}}(f_0, \mathbf{F}^{(k)}) < \pi(\mathbf{F}^{(k)}) + \frac{\nu}{8} . \qquad (14)$$

Choose $0 < \eta = d_0 < 1/9$ so that

$$(1-\eta)^{1/(k-1)} \ge 1 - \frac{1}{f_0} \qquad \text{and} \qquad 4d_0 \log_2 \frac{\mathrm{e}}{3d_0} \le \frac{\nu}{4} \qquad (15)$$

(note that the last inequality uses $x \log_2 x \to 0$ as $x \to 0^+$). For fixed integers f_0 and k and constants $\gamma = 1/2$ and $d_k = d_0$, let

$$\delta_k = \delta_k^{(4.13)}(f_0, k, 1/2, d_0) \qquad (16)$$

be the constant guaranteed by Theorem 4.13. We may assume, without loss of generality, that

$$\delta_k \le d_0 . \qquad (17)$$

For positive integer variables $y_{k-1}, \ldots, y_2$, let

$$\delta(y_{k-1}, \ldots, y_2) = \delta^{(4.13)}(f_0, k, 1/2, d_0, \delta_k, y_{k-1}, \ldots, y_2) \qquad (18)$$
$$r(y_{k-1}, \ldots, y_2) = r^{(4.13)}(f_0, k, 1/2, d_0, \delta_k, y_{k-1}, \ldots, y_2) \qquad (19)$$

be the functions guaranteed by Theorem 4.13.

We now define further constants in terms of the regularity lemma, Theorem 4.12. With input parameters η and δ_k and functions [1] $\delta(y_{k-1}, \ldots, y_2)$ and $r(y_{k-1}, \ldots, y_2)$ defined above, Theorem 4.12 guarantees integer constants

$$t = t^{(4.12)}(\eta, \delta_k, \delta, r) \quad \text{and} \quad n_0 = n^{(4.12)}(\eta, \delta_k, \delta, r). \qquad (20)$$

The constant T mentioned in (13) is set to be

$$T = t! .$$

Now, for $n > n_0$ divisible by T and sufficiently large, we verify (13).

[1] Note that the input functions functions $\delta(y_{k-1}, \ldots, y_2)$ and $r(y_{k-1}, \ldots, y_2)$ have $k-2$ variables while Theorem 4.12 would allow us to consider $k-1$ variables. In particular, Theorem 4.12 would allow us to include a variable y_1 corresponding to the number of vertex classes the output family of partitions $\mathscr{P}$ will have. We have no need for this feature in our argument here, so we hold the variable y_1 constant.

5.4 Proof of (13)

According to Theorem 4.12, every k-graph $\mathcal{G}^{(k)}$ on n vertices (n defined above) admits an $(\eta, \delta(\boldsymbol{a}^{\mathscr{P}}), \boldsymbol{a}^{\mathscr{P}})$-equitable t-bounded family of partitions $\mathscr{P}$ with respect to which $\mathcal{G}^{(k)}$ is $(\delta_k, r(\boldsymbol{a}^{\mathscr{P}}))$-regular. As such, for each $\mathcal{G}^{(k)} \in \mathrm{Forb}(n, \mathbf{F}^{(k)})$, we may associate a family of partitions $\mathscr{P}_{\mathcal{G}^{(k)}}$ (if $\mathcal{G}^{(k)}$ admits multiple such partitions, we simply choose one of them). Accordingly, we may impose an equivalence relation $\sim$ on $\mathrm{Forb}(n, \mathbf{F}^{(k)})$ according to the following rule: for $\mathcal{G}^{(k)}, \widetilde{\mathcal{G}}^{(k)} \in \mathrm{Forb}(n, \mathbf{F}^{(k)})$,

$$\mathcal{G}^{(k)} \sim \widetilde{\mathcal{G}}^{(k)} \iff \mathscr{P}_{\mathcal{G}^{(k)}} = \mathscr{P}_{\widetilde{\mathcal{G}}^{(k)}} . \tag{21}$$

Let $\mathrm{Forb}(n, \mathbf{F}^{(k)}) = \Pi_1 \cup \cdots \cup \Pi_N$ be the partition of $\mathrm{Forb}(n, \mathbf{F}^{(k)})$ induced by $\sim$. To prove (13), we first seek to bound the parameter $N = N(n)$.

Clearly, N is at most the number of t-bounded families of partitions on the vertex set $[n]$. For a fixed vector $\boldsymbol{a} = (a_1, \ldots, a_{k-1})$, there are at most $\prod_{j=1}^{k-1} a_j^{\binom{n}{j}}$ families of partitions $\mathscr{P}(k-1, \boldsymbol{a})$ on the vertex set $[n]$. Consequently,

$$\begin{aligned} N &\le \sum_{\boldsymbol{a}} \Big\{ \prod_{j=1}^{k-1} a_j^{\binom{n}{j}} \colon 1 \le a_j \le t \text{ for } j = 1, \ldots, k-1 \Big\} \\ &\le t^{k-1} \times t^{\sum_{j=1}^{k-1} \binom{n}{j}} = 2^{O(n^{k-1})}. \end{aligned} \tag{22}$$

We now seek to bound $|\Pi_\alpha|$ for every $\alpha = 1, \ldots, N$. Fix $1 \le \alpha \le N$ and, correspondingly, family of partitions $\mathscr{P}_\alpha = \{\mathscr{P}_\alpha^{(1)}, \ldots, \mathscr{P}_\alpha^{(k-1)}\}$, i.e., the family associated to every $\mathcal{G}^{(k)} \in \Pi_\alpha$. With each $\mathcal{G}^{(k)} \in \Pi_\alpha$, we associate the vector

$$\boldsymbol{x}_{\mathcal{G}^{(k)}} = \left(x_{\hat{\mathcal{P}}^{(k-1)}} \colon \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}_\alpha^{(k-1)} \right) \in \{0,1\}^{|\hat{\mathscr{P}}_\alpha^{(k-1)}|} , \tag{23}$$

where, for fixed $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}_\alpha^{(k-1)}$,

$$x_{\hat{\mathcal{P}}^{(k-1)}} = \begin{cases} 1 \text{ if } d(\mathcal{G}^{(k)} | \hat{\mathcal{P}}^{(k-1)}) \ge d_0 \text{ and} \\ \qquad\qquad \mathcal{G}^{(k)} \text{ is } (\delta_k, *, r(\boldsymbol{a}^{\mathscr{P}_\alpha}))\text{-regular w.r.t. } \hat{\mathcal{P}}^{(k-1)}, \\ 0 \text{ otherwise.} \end{cases} \tag{24}$$

From (7) and the t-boundedness of the family $\mathscr{P}_\alpha$,

$$|\{\boldsymbol{x}_{\mathcal{G}^{(k)}} \colon \mathcal{G}^{(k)} \in \Pi_\alpha\}| \le 2^{t^{2^k}} = O(1) . \tag{25}$$

With $\alpha \in [N]$ fixed, fix vector $\boldsymbol{x} \in \{0,1\}^{|\hat{\mathscr{P}}_\alpha^{(k-1)}|}$ and define

$$\Pi_\alpha(\boldsymbol{x}) = \left\{ \mathcal{G}^{(k)} \in \Pi_\alpha \colon \boldsymbol{x}_{\mathcal{G}^{(k)}} = \boldsymbol{x} \right\} .$$

We prove the following lemma.

Lemma 5.2. $\log_2 |\Pi_\alpha(\boldsymbol{x})| \le \mathrm{ex}(n, \mathbf{F}^{(k)}) + \frac{\nu}{2}\binom{n}{k}$.

Lemma 5.2, combined with (22) and (25), easily implies (13) (and hence, Theorem 2.3). Indeed

$$
\begin{aligned}
|\mathrm{Forb}(n, \mathbf{F}^{(k)})| &= \sum_{\alpha=1}^{N} |\Pi_\alpha| = \sum_{\alpha=1}^{N} \sum_{\boldsymbol{x}} |\Pi_\alpha(\boldsymbol{x})| \\
&\le 2^{O(n^{k-1})} \times O(1) \times 2^{\mathrm{ex}(n,\mathbf{F}^{(k)}) + \frac{\nu}{2}\binom{n}{k}} \le 2^{\mathrm{ex}(n,\mathbf{F}^{(k)}) + \nu\binom{n}{k}}
\end{aligned}
$$

where the last inequality holds for sufficiently large n.

We now proceed to prove Lemma 5.2.

5.5 Proof of Lemma 5.2

Fix $\alpha \in [N]$ and, correspondingly, $\mathscr{P}_\alpha = \mathscr{P}_\alpha(k-1, \boldsymbol{a}^{\mathscr{P}_\alpha})$ with $\boldsymbol{a}^{\mathscr{P}_\alpha} = (a_1, \dots, a_{k-1})$ and fix $\boldsymbol{x} = (x_{\hat{\mathcal{P}}^{(k-1)}} \colon \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}_\alpha^{(k-1)})$. Define $\mathcal{D}_\alpha(\boldsymbol{x})$ to be the set of k-tuples $K \in \mathrm{Cross}_k(\mathscr{P}_\alpha^{(1)})$ for which each $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$ is 'dense and regular' w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$:

$$
\mathcal{D}_\alpha(\boldsymbol{x}) = \bigcup \left\{ \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}) \colon x_{\hat{\mathcal{P}}^{(k-1)}} = 1 \text{ (cf. (24))} \right\}. \tag{26}
$$

We make the following claim.

Claim 5.3. $|\mathcal{D}_\alpha(\boldsymbol{x})| \le (\widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}) + \frac{\nu}{4})\binom{n}{k}$.

Our proof of Claim 5.3 is based on the counting lemma, Theorem 4.13. On the other hand, Lemma 5.2 is a simple consequence of Claim 5.3. As such, we go ahead and assume Claim 5.3, for the moment, and finish the proof of Lemma 5.2, before we verify Claim 5.3.

Finishing the proof of Lemma 5.2, note that every edge $K \in \binom{[n]}{k} \setminus \mathcal{D}_\alpha(\boldsymbol{x})$ satisfies that either

(I) K is non-crossing in $\mathscr{P}^{(1)}$,
(II) or $x_{\hat{\mathcal{P}}^{(k-1)}(K)} = 0$, i.e., by (24), polyad $\hat{\mathcal{P}}^{(k-1)}(K)$ is either 'sparse' or 'irregular' (for every $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$).

However, since every $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$ is $(\delta_k, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $(\eta, \delta(\boldsymbol{a}^{\mathscr{P}_\alpha}), \boldsymbol{a}^{\mathscr{P}_\alpha})$-equitable family $\mathscr{P}_\alpha$, the number of edges $K \in \binom{[n]}{k}$ satisfying (I) or (II) is at most

$$
(\eta + \delta_k + d_0)\binom{n}{k} \overset{(15),(17)}{\le} 3d_0\binom{n}{k}.
$$

(Indeed, the equitability of family $\mathscr{P}_\alpha$ ensures that there are at most $\eta\binom{n}{k}$ non-crossing edges. The fact that every $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$ is $(\delta_k, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. family $\mathscr{P}_\alpha$ ensures that at most $\delta_k\binom{n}{k}$ many k-tuples belong to irregular polyads. Finally, sparse polyads (with density smaller than d_0), in total, can only give rise to at most $d_0\binom{n}{k}$ many k-tuples.)

Now, every $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$ can be written as a disjoint union $\mathcal{G}^{(k)} = \mathcal{G}_1^{(k)} \cup \mathcal{G}_2^{(k)}$ where $\mathcal{G}_1^{(k)} \subseteq \mathcal{D}_\alpha(\boldsymbol{x})$ and $|\mathcal{G}_2^{(k)}| \le 3d_0\binom{n}{k}$. As such,

$$|\Pi_\alpha(\boldsymbol{x})| \le 2^{|\mathcal{D}_\alpha(\boldsymbol{x})|} \times \sum_{j=0}^{3d_0\binom{n}{k}} \binom{\binom{n}{k}}{j} \overset{\text{Claim 5.3}}{\le} 2^{(\widetilde{\text{ex}}(n,\mathbf{F}^{(k)})+\frac{\nu}{4})\binom{n}{k}} \times n^k \left(\frac{\mathrm{e}}{3d_0}\right)^{3d_0\binom{n}{k}},$$

which implies (with n large)

$$\log_2 \left|\Pi_\alpha(\boldsymbol{x})\right| \le \left(\widetilde{\text{ex}}(n,\mathbf{F}^{(k)}) + \frac{\nu}{4} + 4d_0 \log \frac{\mathrm{e}}{3d_0}\right)\binom{n}{k} \overset{(15)}{\le} \text{ex}(n,\mathbf{F}^{(k)}) + \frac{\nu}{2}\binom{n}{k},$$

as promised by Lemma 5.2.

It now only remains to prove Claim 5.3.

5.6 Proof of Claim 5.3

Let $\alpha \in [N]$ and $\boldsymbol{x} \in \{0,1\}^{|\hat{\mathscr{P}}_\alpha^{(k-1)}|}$ be fixed. For crossing set $A \in \text{Cross}_{a_1}(\mathscr{P}_\alpha^{(1)})$, define auxiliary k-graph

$$\text{Dense}(A) = \left\{K \in \binom{A}{k} : x_{\hat{\mathcal{P}}^{(k-1)}(K)} = 1 \text{ (cf. (24))}\right\}.$$

Double-counting pairs (A, K) where $K \in \text{Dense}(A)$ and $A \in \text{Cross}_{a_1}(\mathscr{P}_\alpha^{(1)})$ yields

$$\left|\mathcal{D}_\alpha(\boldsymbol{x})\right| \left(\frac{n}{a_1}\right)^{a_1-k} = \sum_{A\in\text{Cross}_{a_1}(\mathscr{P}_\alpha^{(1)})} |\,\text{Dense}(A)|. \tag{27}$$

As such, we may infer Claim 5.3 from the the following assertion:

$$\max\{|\,\text{Dense}(A)| : \; A \in \text{Cross}_{a_1}(\mathscr{P}_\alpha^{(1)})\} < \left(\widetilde{\text{ex}}(a_1,\mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{a_1}{k}. \tag{28}$$

Indeed, since $|\text{Cross}_{a_1}(\mathscr{P}_\alpha^{(1)})| = (n/a_1)^{a_1}$, we combine (27) and (28) to say

$$\left|\mathcal{D}_\alpha(\boldsymbol{x})\right| < \left(\widetilde{\text{ex}}(a_1,\mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{a_1}{k}\left(\frac{n}{a_1}\right)^k \le \left(\widetilde{\text{ex}}(a_1,\mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{n}{k}.$$

Since[2] $a_1 \geq f_0$ (where f_0 is given in (14)) and the sequence $(\widetilde{\mathrm{ex}}(s, \mathbf{F}^{(k)}))_{s=1}^{\infty}$ is non-increasing with limit $\pi(\mathbf{F}^{(k)})$ (see (12)), we have

$$\left|\mathcal{D}_\alpha(\boldsymbol{x})\right| < \left(\widetilde{\mathrm{ex}}(f_0, \mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{n}{k} \overset{(14)}{<} \left(\pi(\mathbf{F}^{(k)}) + \frac{\nu}{4}\right)\binom{n}{k}$$
$$\leq \left(\widetilde{\mathrm{ex}}(n, \mathbf{F}^{(k)}) + \frac{\nu}{4}\right)\binom{n}{k}.$$

Thus, it remains to prove the assertion in (28).

Proof of (28). On the contrary, suppose there exists $A \in \mathrm{Cross}_{a_1}(\mathscr{P}_\alpha^{(1)})$ so that

$$|\,\mathrm{Dense}(A)| \geq \left(\widetilde{\mathrm{ex}}(a_1, \mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{a_1}{k}. \tag{29}$$

As such, we claim there must also exist $B \in \binom{A}{f_0}$ (see (14)) such that the sub-hypergraph $\mathrm{Dense}_B(A)$ of $\mathrm{Dense}(A)$ induced on B contains at least $\mathrm{ex}(f_0, \mathbf{F}^{(k)}) + 1$ edges. Indeed, supposing otherwise, the number M of pairs (K, B), $K \in \binom{B}{k}$, $B \in \binom{A}{f_0}$, would, on the one hand, satisfy

$$M \leq \binom{a_1}{f_0}\mathrm{ex}(f_0, \mathbf{F}^{(k)}) = \widetilde{\mathrm{ex}}(f_0, \mathbf{F}^{(k)})\binom{f_0}{k}\binom{a_1}{f_0}. \tag{30}$$

On the other hand, by the choice of A in (29),

$$M \geq \left(\widetilde{\mathrm{ex}}(a_1, \mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{a_1}{k}\binom{a_1 - k}{f_0 - k}.$$

The monotonicity of the sequence $(\widetilde{\mathrm{ex}}(s, \mathbf{F}^{(k)}) : s \geq 1)$ then gives

$$M \geq \left(\widetilde{\mathrm{ex}}(a_1, \mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{a_1}{k}\binom{a_1 - k}{f_0 - k}$$
$$\geq \left(\pi(\mathbf{F}^{(k)}) + \frac{\nu}{8}\right)\binom{a_1}{k}\binom{a_1 - k}{f_0 - k} \overset{(14)}{>} \widetilde{\mathrm{ex}}(f_0, \mathbf{F}^{(k)})\binom{a_1}{k}\binom{a_1 - k}{f_0 - k},$$

contradicting (30).

Fix $B \in \binom{A}{f_0}$ for which the f_0-vertex sub-hypergraph $\mathrm{Dense}_B(A)$ of $\mathrm{Dense}(A)$ induced on B contains at least $\mathrm{ex}(f_0, \mathbf{F}^{(k)}) + 1$ edges. Then, there

[2] It is easy to see $a_1 \geq f_0$. Indeed, since $\mathscr{P}_\alpha$ is an $(\eta, \delta(\boldsymbol{a}^{\mathscr{P}_\alpha}), \boldsymbol{a}^{\mathscr{P}_\alpha})$-equitable family of partitions and since $|\mathrm{Cross}_k(\mathscr{P}_\alpha^{(1)})| = \binom{a_1}{k}(\frac{n}{a_1})^k$, we have

$$1 - \eta \leq |\mathrm{Cross}_k(\mathscr{P}_\alpha^{(1)})|\binom{n}{k}^{-1} \leq \left(1 - \frac{1}{a_1}\right)^{k-1}$$

where the last inequality holds with n sufficiently large. The assertion $a_1 \geq f_0$ then follows from our choice of η in (15), i.e., $1 - \eta \geq (1 - f_0^{-1})^{k-1}$.

exists $\mathcal{F}^{(k)} \in \mathbf{F}^{(k)}$ so that its copy $\mathcal{F}_0^{(k)}$ appears as a sub-hypergraph of $\mathrm{Dense}_B(A)$. In order to derive a contradiction from our assumption in (29), we use the counting lemma, Theorem 4.13, to find a copy of the same $\mathcal{F}^{(k)} \in \mathbf{F}^{(k)}$ in any (and every) $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$. Since $\Pi_\alpha(\boldsymbol{x}) \subseteq \mathrm{Forb}(\mathbf{F}^{(k)})$, we have an immediate contradiction.

To that end, fix $\mathcal{G}^{(k)} \in \Pi_\alpha(\boldsymbol{x})$ and let $F = V(\mathcal{F}_0^{(k)}) \subseteq B$. For each $K \in \binom{F}{k}$, set

$$\mathcal{H}_K^{(k)} = \begin{cases} \mathcal{G}^{(k)} \cap \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K)) & \text{if } \quad K \in \mathcal{F}_0^{(k)}, \\ \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K)) & \text{otherwise.} \end{cases}$$

Set

$$\mathcal{H}^{(k)} = \bigcup \left\{ \mathcal{H}_K^{(k)} \colon K \in \binom{F}{k} \right\}.$$

With $\mathcal{H}^{(k)}$ defined above, observe that every element of $\mathcal{K}_f(\mathcal{H}^{(k)})$, $f = |F|$ corresponds to a copy of $\mathcal{F}^{(k)}$ appearing as a sub-hypergraph of $\mathcal{G}^{(k)}$. If we show $|\mathcal{K}_f(\mathcal{H}^{(k)})| > 0$, then we derive a contradiction, and hence, (28) follows.

To show $|\mathcal{K}_f(\mathcal{H}^{(k)})| > 0$, we apply the counting lemma, Theorem 4.13, to $\mathcal{H}^{(k)}$ and $\boldsymbol{\mathcal{Q}} = \{\mathcal{Q}^{(j)}\}_{j=1}^{k-1}$ where $\mathcal{Q}^{(j)} = \bigcup \left\{ \mathcal{P}^{(j)}(J) \colon J \in \binom{F}{j} \right\}$ for $j = 1, \ldots, k-1$. We first check that the assumptions of Theorem 4.13 are met by $\mathcal{H}^{(k)}$ and $\boldsymbol{\mathcal{Q}}$:

1. Since $\mathscr{P}_\alpha$ is an $(\eta, \delta(\boldsymbol{a}^{\mathscr{P}}), \boldsymbol{a}^{\mathscr{P}})$-equitable family, the $(n/a_1, f, k-1)$-complex $\boldsymbol{\mathcal{Q}}$ is $(\delta(\boldsymbol{a}^{\mathscr{P}}), (1/a_2, \ldots, 1/a_{k-1}))$-regular. Moreover, we chose the function δ in (18) appropriately for an application of Theorem 4.13;
2. For each $K \in \mathcal{F}_0^{(k)} \subseteq \mathrm{Dense}_B(A) \subseteq \mathrm{Dense}(A)$, the definition of $\boldsymbol{x}$ in (23) guarantees that $\mathcal{H}_K^{(k)} = \mathcal{G}^{(k)} \cap \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K))$ is $(\delta_k, *, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K) \subseteq \mathcal{Q}^{(k-1)}$ and that $d(\mathcal{G}^{(k)} | \hat{\mathcal{P}}^{(k-1)}(K)) \geq d_0$. We note that δ_k and r were chosen in (16) and (19) appropriately for an application of Theorem 4.13;
3. For each $K \in \binom{F}{k} \setminus \mathcal{F}_0^{(k)}$, the k-graph $\mathcal{H}_K^{(k)} = \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K))$ is easily seen to be $(\varepsilon, 1, s)$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$ for every $\varepsilon > 0$ and $s \in \mathbb{N}$. As such, $\mathcal{H}_K$ is $(\delta_k, 1, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$.

Hence, we can apply the hypergraph counting lemma to $\mathcal{H}^{(k)}$ and $\boldsymbol{\mathcal{Q}}$. We conclude

$$|\mathcal{K}_f(\mathcal{H}^{(k)})| \geq \frac{1}{2} d_0^{\binom{f}{k}} \prod_{j=2}^{k-1} \left(\frac{1}{a_j}\right)^{\binom{f}{j}} \left(\frac{n}{a_1}\right)^f \geq \frac{1}{2} d_0^{\binom{f_0}{k}} \prod_{j=2}^{k-1} \left(\frac{1}{a_j}\right)^{\binom{f_0}{j}} \left(\frac{n}{a_1}\right)^{f_0} > 0.$$

This proves (28). □

6 Proof of Theorem 2.4

Theorem 2.4 is a simple consequence of the following lemma.

Lemma 6.1. *Let k-graph $\mathcal{F}^{(k)}$ on f vertices be given. For every $c > 0$, there exist $\varepsilon > 0$ and integers $\widetilde{r}$, T and n_0 so that a given k-graph $\mathcal{G}^{(k)}$ on vertex set $[n] = \{1, \dots, n\}$, with $n \geq n_0$ and n divisible by T, is an induced Ramsey k-graph for $\mathcal{F}^{(k)}$ whenever the following conditions are met:*

(i) $|\mathcal{K}_s(\mathcal{G}^{(k)})| \geq c\binom{n}{s}$ *where* $s = R^{(k)}(f, f)$ *is the Ramsey number for* $K_f^{(k)}$*;*
(ii) $\mathcal{G}^{(k)}$ *is* $(\varepsilon, d(\mathcal{G}^{(k)}|\mathcal{P}^{(k-1)}), \widetilde{r})$*-regular,* $d(\mathcal{G}^{(k)}|\mathcal{P}^{(k-1)}) \in [\frac{1}{4}, \frac{3}{4}]$*, w.r.t. every* $(k-1)$*-graph* $\mathcal{P}^{(k-1)} \subseteq \binom{[n]}{k-1}$ *which satisfies* $|\mathcal{K}_k(\mathcal{P}^{(k-1)})| \geq n^k/\log n$*.*

Lemma 6.1 implies Theorem 2.4. Indeed, it is easy to verify that, with probability tending to 1 as $n \to \infty$, i.e., *asymptotically almost surely (a.a.s.)*, the binomial random k-graph $\mathcal{G}^{(k)}(n, 1/2)$ satisfies the hypothesis of Lemma 6.1 with $c = (1/2)^{\binom{s}{k}-1}$ and with arbitrary choices of $\varepsilon > 0$ and integers $\widetilde{r}$ and T. In particular, Chebyshev's inequality verifies that $\mathcal{G}^{(k)}(n, 1/2)$ satisfies *(i)*, a.a.s. For completeness, we verify in the Appendix (see Fact A.5) that $\mathcal{G}^{(k)}(n, 1/2)$ satisfies a.a.s. *(ii)*.

The goal of this section is, therefore, to prove Lemma 6.1. As our proof depends on Theorems 4.12 and 4.13, we again first discuss a sequence of auxiliary constants.

6.1 Constants

Let k-graph $\mathcal{F}^{(k)}$ on f vertices be given. Set, as in the hypothesis of Lemma 6.1,

$$s = R^{(k)}(f, f). \tag{31}$$

As in the hypothesis of Lemma 6.1, let $c > 0$ be given. We define $\varepsilon > 0$ and integers $\widetilde{r}$ and t in terms of Theorem 4.12 and 4.13.

As in Theorem 4.13, put $\ell = f$, $\gamma = 1/2$ and $d_k = 1/8$ and let

$$\delta_k^{(4.13)} = \delta_k^{(4.13)}(f, k, 1/2, d_k)$$

be the constant guaranteed by Theorem 4.13. Set

$$\eta = \delta_k = \min\left\{\frac{1}{2}\delta_k^{(4.13)}, \frac{c}{4}\binom{s}{k}^{-1}\right\} \tag{32}$$

For positive integer variables $y_{k-1}, \dots, y_2$, let

$$\delta(y_{k-1}, \dots, y_2) = \delta^{(4.13)}(f, k, 1/2, d_k, y_{k-1}, \dots, y_2), \tag{33}$$

$$r(y_{k-1}, \dots, y_2) = r^{(4.13)}(f, k, 1/2, d_k, y_{k-1}, \dots, y_2) \tag{34}$$

be the functions guaranteed by Theorem 4.13. Without loss of generality, we assume that $r(y_{k-1},\dots,y_2)$ is monotone increasing in every coordinate.

We now define more auxiliary constants. In Theorem 4.12, let constants η and δ_k and functions r and δ be the parameters chosen in (32)–(34). Theorem 4.12 guarantees integer *constants*

$$t = t^{(4.12)}(\eta,\delta_k,r,\delta) \quad \text{and} \quad n_0 = n_0^{(4.12)}(\eta,\delta_k,r,\delta). \tag{35}$$

We set

$$\varepsilon = \delta_k, \quad T = t! \quad \text{and} \quad \widetilde{r} = r(t,\dots,t). \tag{36}$$

Let $n > n_0$ be divisible by T and be sufficiently large whenever needed. This concludes our discussion of the constants.

6.2 Proof of Lemma 6.1

With the constants above, let $\mathcal{G}^{(k)}$ be a k-graph on n vertices satisfying the hypothesis of Lemma 6.1. Let $\mathcal{G}^{(k)} = \mathcal{R}^{(k)} \cup \mathcal{B}^{(k)}$ be any two-coloring with colors 'red' and 'blue'. We prove that one of $\mathcal{R}^{(k)}$ or $\mathcal{B}^{(k)}$ contains a copy of $\mathcal{F}^{(k)}$ as a sub-hypergraph which is induced in $\mathcal{G}^{(k)}$.

With constants η, and δ_k and functions r and δ defined above, we apply Theorem 4.12 to the k-graph $\mathcal{R}^{(k)}$ to obtain $(\eta,\delta(\boldsymbol{a}^{\mathscr{P}}),\boldsymbol{a}^{\mathscr{P}})$-equitable and t-bounded family of partitions $\mathscr{P} = \mathscr{P}(k-1,\boldsymbol{a}^{\mathscr{P}})$ with respect to which $\mathcal{R}^{(k)}$ is $(\delta_k, r(\boldsymbol{a}^{\mathscr{P}}))$-regular. Observe, that due to our choice of $\widetilde{r}$ in (36) and the monotonicity thereof,

$$r(\boldsymbol{a}^{\mathscr{P}}) \le \widetilde{r}. \tag{37}$$

We now consider the polyads of $\mathscr{P}$.

Set[3]

$$\hat{\mathscr{P}}^{(k-1)}_{\text{bad}} = \left\{ \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)} \colon |\mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)})| < n^k/\log n \right\}.$$

Note that the t-boundedness of $\mathscr{P}$ gives for sufficiently large n

$$\begin{aligned}\left| \bigcup \left\{ \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}) \colon \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\text{bad}} \right\} \right| &\le \binom{a_1}{k} \prod_{j=2}^{k-1} a_j^{\binom{k}{j}} \times \frac{n^k}{\log n} \\ &\le \frac{c}{4\binom{s}{k}} \binom{n}{k}. \end{aligned} \tag{38}$$

[3] We note that one could, in fact, show that $\hat{\mathscr{P}}^{(k-1)}_{\text{bad}} = \emptyset$. This would follow from the fact that there are only a bounded number (independent of n) of polyads $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}$ and each of them corresponds to a $(\delta,(1/a_2,\dots,1/a_{k-1}))$-regular $(n/a_1,k,k-1)$-complex. In this situation, one can argue that with $\delta \ll \min\{1/a_1,\dots,1/a_{k-1}\}$ we have $|\mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)})| = (1 \pm f(\delta))\prod_{h=2}^{k-1}(1/a_h)^{\binom{k}{h}} \times (n/a_1)^k$ where $f(\delta) \to 0$ as $\delta \to 0$. Rather than making this precise, however, we chose in our current proof to use the fact that (sparse) polyads $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\text{bad}}$ can have only little influence.

Set

$$\hat{\mathscr{P}}^{(k-1)}_{\text{reg}} = \Big\{ \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)} \setminus \hat{\mathscr{P}}^{(k-1)}_{\text{bad}} : \mathcal{R}^{(k)} \text{ is } (\delta_k, *, r(\boldsymbol{a}^{\mathscr{P}}))\text{-regular w.r.t. } \hat{\mathcal{P}}^{(k-1)} \Big\}. \tag{39}$$

While $\hat{\mathscr{P}}^{(k-1)}_{\text{reg}}$ is defined in terms of the k-graph $\mathcal{R}^{(k)}$ only, the following fact observes that both $\mathcal{R}^{(k)}$ and $\mathcal{B}^{(k)}$ are 'regular' w.r.t. every polyad $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\text{reg}}$.

Fact 6.2.

$$\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\text{reg}} \quad \Longrightarrow \quad \mathcal{B}^{(k)} \textit{ is } (2\delta_k, *, r(\boldsymbol{a}^{\mathscr{P}}))\textit{-regular w.r.t. } \hat{\mathcal{P}}^{(k-1)}.$$

Proof of Fact 6.2. Indeed, for fixed $\hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\text{reg}}$, we know

1. $\mathcal{R}^{(k)}$ is $(\delta_k, *, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}$ (by definition of $\hat{\mathscr{P}}^{(k-1)}_{\text{reg}}$);
2. $\mathcal{G}^{(k)}$ is $(\varepsilon, d(\mathcal{G}^{(k)}|\hat{\mathcal{P}}^{(k-1)}), \widetilde{r})$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}$, where $d(\mathcal{G}^{(k)}|\hat{\mathcal{P}}^{(k-1)}) \in [\frac{1}{4}, \frac{3}{4}]$ (see (ii) of Lemma 6.1).

As such, it may be directly verified from Definition 4.10 that the difference $\mathcal{B}^{(k)} = \mathcal{G}^{(k)} \setminus \mathcal{R}^{(k)}$ is $(\varepsilon + \delta_k, *, \min\{r(\boldsymbol{a}^{\mathscr{P}}), \widetilde{r}\})$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}$ (with complementary density $d(\mathcal{B}^{(k)}|\hat{\mathcal{P}}^{(k-1)}) = d(\mathcal{G}^{(k)}|\hat{\mathcal{P}}^{(k-1)}) - d(\mathcal{R}^{(k)}|\hat{\mathcal{P}}^{(k-1)})$). Recalling $\varepsilon = \delta_k$ from (36) and $r(\boldsymbol{a}^{\mathscr{P}}) \leq \widetilde{r}$ from (37), Fact 6.2 follows. ◻

We proceed with the first of two easy claims that will prove Lemma 6.1.

Claim 6.3. For $s = R^{(k)}(f, f)$ fixed in (31), there exists $S \in \text{Cross}_s(\mathscr{P}^{(1)})$ so that every $K \in \binom{S}{k}$ has $\hat{\mathcal{P}}^{(k-1)}(K) \in \hat{\mathscr{P}}^{(k-1)}_{\text{reg}}$.

Proof of Claim 6.3. Set

$$\widetilde{\mathcal{G}}^{(k)} = \mathcal{G}^{(k)} \cap \text{Cross}_k(\mathscr{P}^{(1)}) \cap \Big\{ \bigcup \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}) : \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\text{reg}} \Big\}. \tag{40}$$

Observe that every $S \in \mathcal{K}_s(\widetilde{\mathcal{G}}^{(k)})$ satisfies the properties required by the claim. As such, it suffices to prove $|\mathcal{K}_s(\widetilde{\mathcal{G}}^{(k)})| > 0$. Recall that our hypothesis in Lemma 6.1 assumes that $|\mathcal{K}_s(\mathcal{G}^{(k)})| > c\binom{n}{s}$. We show that, in deleting the few edges of $\mathcal{G}^{(k)}$ to obtain $\widetilde{\mathcal{G}}^{(k)}$, we don't destroy all of these cliques.

First, we check that $|\mathcal{G}^{(k)} \setminus \widetilde{\mathcal{G}}^{(k)}|$ is small. Indeed, since $\mathscr{P}$ is an $(\eta, \delta(\boldsymbol{a}^{\mathscr{P}}), \boldsymbol{a}^{\mathscr{P}})$-equitable family of partitions,

$$\left| \mathcal{G}^{(k)} \setminus \text{Cross}_k(\mathscr{P}^{(1)}) \right| \leq \eta \binom{n}{k}. \tag{41}$$

Combining (38) with the fact that $\mathcal{R}^{(k)}$ is $(\delta_k, r(a^{\mathscr{P}}))$-regular w.r.t. $\mathscr{P}$ we have (in view of (39)) that

$$\Big|\mathcal{G}^{(k)} \setminus \Big\{ \bigcup \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)})\colon \hat{\mathcal{P}}^{(k-1)} \in \hat{\mathscr{P}}^{(k-1)}_{\mathrm{reg}} \Big\}\Big| \le \Big(\delta_k + \frac{c}{4\binom{s}{k}}\Big)\binom{n}{k}. \tag{42}$$

Consequently, we infer from (40), (41), and (42) that

$$|\mathcal{G}^{(k)} \setminus \widetilde{\mathcal{G}}^{(k)}| \le \Big(\eta + \delta_k + \frac{c}{4\binom{s}{k}}\Big)\binom{n}{k} \overset{(32)}{\le} \frac{3c}{4}\binom{s}{k}^{-1}\binom{n}{k}. \tag{43}$$

Now, since each k-tuple of $\mathcal{G}^{(k)} \setminus \widetilde{\mathcal{G}}^{(k)}$ can belong to at most $\binom{n-k}{s-k}$ cliques $K_s^{(k)}$, we see that (43) implies

$$\begin{aligned}\big|\mathcal{K}_s(\widetilde{\mathcal{G}}^{(k)})\big| &\ge \big|\mathcal{K}_s(\mathcal{G}^{(k)})\big| - \frac{3c}{4}\binom{s}{k}^{-1}\binom{n}{k}\binom{n-k}{s-k}\\ &= \big|\mathcal{K}_s(\mathcal{G}^{(k)})\big| - \frac{3c}{4}\binom{n}{s} \overset{(i)}{\ge} \frac{c}{4}\binom{n}{s} > 0\end{aligned}$$

where we used property (i) from the hypothesis of Lemma 6.1. □

As guaranteed by Claim 6.3, fix $S \in \mathrm{Cross}_s(\mathscr{P}^{(1)})$ of size $s = R^{(k)}(f,f)$ whose every $K \in \binom{S}{k}$ has $\hat{\mathcal{P}}^{(k-1)}(K) \in \hat{\mathscr{P}}^{(k-1)}_{\mathrm{reg}}$. We continue with the second of two easy claims that will prove Lemma 6.1.

Claim 6.4. There exists a set $F \in \binom{S}{f}$ such that either

$$d(\mathcal{R}^{(k)}|\hat{\mathcal{P}}^{(k-1)}(K)) \ge \frac{1}{2} \quad \text{for every } K \in \binom{F}{k}, \tag{44}$$

or

$$d(\mathcal{B}^{(k)}|\hat{\mathcal{P}}^{(k-1)}(K)) \ge \frac{1}{2} \quad \text{for every } K \in \binom{F}{k}. \tag{45}$$

Proof of Claim 6.4. For each $K \in \binom{S}{k}$, define an auxiliary two-coloring

$$\chi(K) = \begin{cases} \text{'red' if } & d(\mathcal{R}^{(k)}|\hat{\mathcal{P}}^{(k-1)}(K)) \ge \frac{1}{2}, \\ \text{'blue' otherwise.} \end{cases}$$

Note that, by the definition of $\hat{\mathscr{P}}^{(k-1)}_{\mathrm{bad}}$ and assumption (ii) of Lemma 6.1, we have $d(\mathcal{G}^{(k)}|\hat{\mathcal{P}}^{(k-1)}) \ge 1/4$ for every $\hat{\mathcal{P}}^{(k-1)} \notin \hat{\mathscr{P}}^{(k-1)}_{\mathrm{bad}}$. Consequently, for every $\hat{\mathcal{P}}^{(k-1)} \notin \hat{\mathscr{P}}^{(k-1)}_{\mathrm{bad}}$ either $d(\mathcal{R}^{(k)}|\hat{\mathcal{P}}^{(k-1)}) \ge 1/8$ or $d(\mathcal{B}^{(k)}|\hat{\mathcal{P}}^{(k-1)}) \ge 1/8$. (In this way, $\chi(K) =$'blue' implies $d(\mathcal{B}^{(k)}|\hat{\mathcal{P}}^{(k-1)}(K)) \ge 1/8$.) Now, it follows from $s = R^{(k)}(f,f)$ in (31) that there exists a set $F \in \binom{S}{f}$ such that χ is constant on $\binom{F}{k}$. Claim 6.4 then follows. □

We now deduce Lemma 6.1 from Claim 6.3 and 6.4. Let $F \in \binom{S}{f}$ be the set with the properties guaranteed by Claim 6.4. Then, the polyads $\hat{\mathcal{P}}^{(k-1)}(K)$

across $K \in \binom{F}{k}$ are all 'dense' in the same color $\mathcal{R}^{(k)}$ or $\mathcal{B}^{(k)}$. Recall Fact 6.2 ensures these same polyads are also all 'regular' across $K \in \binom{F}{k}$. It therefore doesn't matter which of (44) or (45) holds, and so we assume, without loss of generality, that the former does.

Fix a copy $\mathcal{F}_0^{(k)}$ of $\mathcal{F}^{(k)}$ on the set F, i.e., $V(\mathcal{F}_0^{(k)}) = F$. We construct a sub-hypergraph $\mathcal{H}^{(k)} \subseteq \mathcal{G}^{(k)}$ as follows. For $K \in \binom{F}{k} = \binom{V(\mathcal{F}_0^{(k)})}{k}$, set

$$\mathcal{H}_K^{(k)} = \begin{cases} \mathcal{R}^{(k)} \cap \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K)) \text{ if } \quad K \in \mathcal{F}_0^{(k)}, \\ \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K)) \setminus \mathcal{G}^{(k)} \quad \text{otherwise.} \end{cases}$$

Define

$$\mathcal{H}^{(k)} = \bigcup \left\{ \mathcal{H}_K^{(k)} \colon K \in \binom{F}{k} \right\}.$$

With $\mathcal{H}^{(k)}$ defined above, observe that every element of $\mathcal{K}_f(\mathcal{H}^{(k)})$ corresponds to a copy of $\mathcal{F}^{(k)} \subset \mathcal{R}^{(k)}$ which is induced in $\mathcal{G}^{(k)}$. To conclude the proof of Lemma 6.1, therefore, it suffices to show $|\mathcal{K}_f(\mathcal{H}^{(k)})| > 0$. To this end, we use the counting lemma, Theorem 4.13, and first check that it is appropriate to do so.

Indeed, for $j = 1, \ldots, k-1$, set

$$\mathcal{Q}^{(j)} = \bigcup \left\{ \mathcal{P}^{(j)}(J) \colon J \in \binom{F}{j} \right\}$$

and $\boldsymbol{\mathcal{Q}} = \{\mathcal{Q}^{(j)}\}_{j=1}^{k-1}$. We observe the following.

1. $\boldsymbol{\mathcal{Q}}$ is a $(\delta(\boldsymbol{a}^{\mathscr{P}}), r(1/a_2, \ldots, 1/a_{k-1}))$-regular $(n/a_1, f, k-1)$-complex, where the function δ was chosen in (33) appropriately for an application of Theorem 4.13;
2. For $K \in \mathcal{F}_0^{(k)}$, we combine Claim 6.3 and Claim 6.4 to see that $\mathcal{H}_K^{(k)} = \mathcal{R}^{(k)} \cap \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K))$ is $(\delta_k, *, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$ with density $d(\mathcal{R}^{(k)} | \hat{\mathcal{P}}^{(k-1)}(K)) \geq \frac{1}{8}$. We note that $\delta_k \leq \delta_k^{(4.13)}(f, k, 1/2, d_k)$ and $r = r(\boldsymbol{a}^{\mathscr{P}})$ were chosen in (32) and (34), resp., appropriately for an application of Theorem 4.13;
3. For each $K \in \binom{F}{k} \setminus \mathcal{F}_0^{(k)}$, we have, by (ii) of Lemma 6.1, that $\mathcal{G}^{(k)}$ is $(\varepsilon, d_K, \widetilde{r})$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$ with $d_K = d(\mathcal{G}^{(k)} | \hat{\mathcal{P}}^{(k-1)}(K)) \in [\frac{1}{4}, \frac{3}{4}]$. Since $\varepsilon = \delta_k$ and $\widetilde{r} \geq r(\boldsymbol{a}^{\mathscr{P}})$ (cf. (36) and (37)), the k-graph $\mathcal{G}^{(k)}$ is therefore also $(\delta_k, d_K, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$. As such, the complement $\mathcal{H}_K^{(k)} = \mathcal{K}_k(\hat{\mathcal{P}}^{(k-1)}(K)) \setminus \mathcal{G}^{(k)}$ is then also $(\delta_k, d_K, r(\boldsymbol{a}^{\mathscr{P}}))$-regular w.r.t. $\hat{\mathcal{P}}^{(k-1)}(K)$ with density $\bar{d}_K = d(\mathcal{H}_K^{(k)} | \hat{\mathcal{P}}^{(k-1)}(K)) = 1 - d_K \in [\frac{1}{4}, \frac{3}{4}]$.

Hence, we can apply the counting lemma to $\mathcal{H}^{(k)}$ and $\boldsymbol{\mathcal{Q}}$. As such, we conclude

$$\left| \mathcal{K}_f(\mathcal{H}^{(k)}) \right| \geq \frac{1}{2} \left(\frac{1}{8} \right)^{\binom{f}{k}} \prod_{j=2}^{k-1} \left(\frac{1}{a_j} \right)^{\binom{f}{j}} \left(\frac{n}{a_1} \right)^{f} > 0,$$

and Lemma 6.1 is proved.

Appendix A

Fact A.5. *With probability at least* $(1-\exp(-n^k/\log^6 n))$ *the binomial random hypergraph* $\mathcal{G}^{(k)}(n,1/2)$ *is* $(1/\log n, 1/2, \log n)$*-regular w.r.t. to every* $(k-1)$*-uniform hypergraph* $\mathcal{P}^{(k-1)} \subseteq \binom{[n]}{k-1}$ *for which* $|\mathcal{K}_k(\mathcal{P}^{(k-1)})| \geq n^k/\log n$*.*

Proof. The proof of Fact A.5 follows standard lines. For simplicity of notation, set $r = \log n$. Fix any $(k-1)$-graph $\mathcal{P}^{(k-1)} \subseteq \binom{[n]}{k-1}$ for which

$$\left|\mathcal{K}_k(\mathcal{P}^{(k-1)})\right| > \frac{n^k}{\log n}. \tag{46}$$

Let $\boldsymbol{\mathcal{Q}}^{(k-1)} = \{\mathcal{Q}_1^{(k-1)}, \dots, \mathcal{Q}_r^{(k-1)}\}$ be a family of r sub-hypergraphs of $\mathcal{P}^{(k-1)}$ for which

$$\left|\bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})\right| \geq \frac{1}{\log n}\left|\mathcal{K}_k(\mathcal{P}^{(k-1)})\right| \overset{(46)}{>} \frac{n^k}{\log^2 n}. \tag{47}$$

Set $X(\boldsymbol{\mathcal{Q}}^{(k-1)}) = |\mathcal{G}^{(k)}(n,1/2) \cap \bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})|$. Then, $X(\boldsymbol{\mathcal{Q}}^{(k-1)})$ is binomially distributed random variable with expectation

$$\begin{aligned}\mathbb{E}\big[X(\boldsymbol{\mathcal{Q}}^{(k-1)})\big] &= \mathbb{E}\left[\left|\mathcal{G}^{(k)}(n,1/2) \cap \bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})\right|\right] \\ &= \frac{1}{2}\left|\bigcup_{i\in[r]} \mathcal{K}_k(\mathcal{Q}_i^{(k-1)})\right| \overset{(47)}{>} \frac{n^k}{2\log^2 n}.\end{aligned} \tag{48}$$

We apply the Chernoff inequality (cf. [JŁR00]) to conclude

$$\begin{aligned}&\mathbf{P}\left(\left|X(\boldsymbol{\mathcal{Q}}^{(k-1)}) - \mathbb{E}\big[X(\boldsymbol{\mathcal{Q}}^{(k-1)})\big]\right| > \frac{1}{\log n}\mathbb{E}\big[X(\boldsymbol{\mathcal{Q}}^{(k-1)})\big]\right) \\ &\qquad\qquad \leq 2\exp\left\{-\frac{\mathbb{E}\big[X(\boldsymbol{\mathcal{Q}}^{(k-1)})\big]}{3\log^2 n}\right\} \overset{(48)}{<} 2\exp\left\{-\frac{n^k}{6\log^4 n}\right\}.\end{aligned} \tag{49}$$

For a given $\mathcal{P}^{(k-1)}$ satisfying (46), let $B(\mathcal{P}^{(k-1)})$ be the event that there exist a family $\boldsymbol{\mathcal{Q}}^{(k-1)} = \{\mathcal{Q}_1^{(k-1)}, \dots, \mathcal{Q}_r^{(k-1)}\}$ of r sub-hypergraphs of $\mathcal{P}^{(k-1)}$ such that (47) and $|X(\boldsymbol{\mathcal{Q}}^{(k-1)}) - \mathbb{E}[X(\boldsymbol{\mathcal{Q}}^{(k-1)})]| > \mathbb{E}[X(\boldsymbol{\mathcal{Q}}^{(k-1)})]/\log n$. As there are at most $2^{r|\mathcal{P}^{(k-1)}|} \leq 2^{n^{k-1}\log n}$ families $\boldsymbol{\mathcal{Q}}^{(k-1)}$ of sub-hypergraphs of $\mathcal{P}^{(k-1)}$, we see

$$\mathbf{P}\big(B(\mathcal{P}^{(k-1)})\big) \overset{(49)}{<} 2\cdot 2^{n^{k-1}\log n}\exp\left\{-\frac{n^k}{6\log^4 n}\right\} < \exp\left\{-\frac{n^k}{\log^5 n}\right\}. \tag{50}$$

We now conclude the proof of Fact A.5. Note that (50) almost proves what we want. Namely, we have fixed an appropriate $(k-1)$-graph $\mathcal{P}^{(k-1)}$ (i.e., which satisfies (46)) and have proved that it is very unlikely that $\mathcal{G}^{(k)}(n,1/2)$ fails to be $(1/\log n, 1/2, \log n)$-regular w.r.t. $\mathcal{P}^{(k-1)}$. We simply want the same assertion for every appropriate $(k-1)$-graph $\mathcal{P}^{(k-1)}$, Since there are at most $2^{\binom{n}{k-1}}$ many $(k-1)$-graphs $\mathcal{P}^{(k-1)}$ satisfying (46), we see

$$\mathbb{P}\Big(\bigcup\big\{B(\mathcal{P}^{(k-1)})\colon \mathcal{P}^{(k-1)} \subseteq \binom{[n]}{k-1} \text{ satisfying (46)}\big\}\Big) \overset{(50)}{<} 2^{\binom{n}{k-1}} \exp\left\{-\frac{n^k}{\log^5 n}\right\} < \exp\left\{-\frac{n^k}{\log^6 n}\right\}.$$

This concludes our proof of Fact A.5. □

References

[AH78] F. G. Abramson and L. A. Harrington, Models without indiscernibles, *J. Symbolic Logic* **43** (1978), no. 3, 572–600.

[AS06+] N. Alon and A. Shapira, On an extremal problem of Brown, Erdős and Sós, *Combinatorica*, to appear.

[ARS06+] C. Avart, V. Rödl, and M. Schacht, Every monotone 3-graph property is testable, submitted.

[BT95] B. Bollobás and A. Thomason, Projections of bodies and hereditary properties of hypergraphs, *Bull. London Math. Soc.* **27** (1995), no. 5, 417–424.

[BT97] B. Bollobás and A. Thomason, Hereditary and monotone properties of graphs, *The mathematics of Paul Erdős, II, Algorithms Combin.*, vol. 14, Springer, Berlin, 1997, pp. 70–78.

[BES73] W. G. Brown, P. Erdős, and V. T. Sós, Some extremal problems on r-graphs, *New directions in the theory of graphs (Proc. Third Ann Arbor Conf., Univ. Michigan, Ann Arbor, Mich, 1971)*, Academic Press, New York, 1973, pp. 53–63.

[EFR86] P. Erdős, P. Frankl, and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.* **2** (1986), no. 2, 113–121.

[EK68] P. Erdős and D. J. Kleitman, On coloring graphs to maximize the proportion of multicolored k-edges, *J. Combinatorial Theory* **5** (1968), 164–169.

[EKR76] P. Erdős, D. J. Kleitman, and B. L. Rothschild, Asymptotic enumeration of K_n-free graphs, *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973)*, Tomo II, Accad. Naz. Lincei, Rome, 1976, pp. 19–27. Atti dei Convegni Lincei, No. 17.

[FR02] P. Frankl and V. Rödl, Extremal problems on set systems, *Random Structures Algorithms* **20** (2002), no. 2, 131–164.

[FK78] H. Furstenberg and Y. Katznelson, An ergodic Szemerédi theorem for commuting transformations, *J. Analyse Math.* **34** (1978), 275–291.

[Gow06+] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, submitted.

[Gow06] W. T. Gowers, Quasirandomness, counting and regularity for 3-uniform hypergraphs, *Combin. Probab. Comput.* **15** (2006), no. 1–2, 143–184.

[HNR03] P. E. Haxell, B. Nagle, and V. Rödl, Integer and fractional packings in dense 3-uniform hypergraphs, *Random Structures Algorithms* **22** (2003), no. 3, 248–310.

[JŁR00] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000.

[KNS64] G. Katona and T. Nemetz, M. Simonovits, On a problem of Turán in the theory of graphs, *Mat. Lapok* **15** (1964), 228–238.

[KNR02] Y. Kohayakawa, B. Nagle, and V. Rödl, Efficient testing of hypergraphs (extended abstract), *Automata, languages and programming*, *Lecture Notes in Comput. Sci.*, vol. 2380, Springer, Berlin, 2002, pp. 1017–1028.

[KNR03] Y. Kohayakawa, B. Nagle, and V. Rödl, Hereditary properties of triple systems, *Combin. Probab. Comput.* **12** (2003), no. 2, 155–189.

[KS96] J. Komlós and M. Simonovits, Szemerédi's regularity lemma and its applications in graph theory, *Combinatorics, Paul Erdős is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud.*, vol. 2, János Bolyai Math. Soc., Budapest, 1996, pp. 295–352.

[KSSS02] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi, The regularity lemma and its applications in graph theory, *Theoretical aspects of computer science (Tehran, 2000), Lecture Notes in Comput. Sci.*, vol. 2292, Springer, Berlin, 2002, pp. 84–112.

[NR98] B. Nagle and V. Rödl, The asymptotic number of triple systems not containing a fixed one, *Discrete Math.* **235** (2001), no. 1-3, 271–290, Combinatorics (Prague, 1998).

[NRS06] B. Nagle, V. Rödl, and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures Algorithms* **28** (2006), no. 2, 113–179.

[NR77] J. Nešetřil and V. Rödl, Partitions of finite relational and set systems, *J. Combinatorial Theory Ser. A* **22** (1977), no. 3, 289–312.

[NR82] J. Nešetřil and V. Rödl, Two proofs of the Ramsey property of the class of finite hypergraphs, *European J. Combin.* **3** (1982), no. 4, 347–352.

[PS92] H. J. Prömel and A. Steger, Excluding induced subgraphs. III. A general asymptotic, *Random Structures Algorithms* **3** (1992), no. 1, 19–31.

[Rödl91] V. Rödl, Some developments in Ramsey theory, *Proceedings of the International Congress of Mathematicians, Vol. I, II (Kyoto, 1990) (Tokyo)*, Math. Soc. Japan, 1991, pp. 1455–1466.

[RR98] V. Rödl and A. Ruciński, Ramsey properties of random hypergraphs, *J. Combin. Theory Ser. A* **81** (1998), no. 1, 1–33.

[RRS06] V. Rödl, A. Ruciński, and E. Szemerédi, Dirac's theorem for 3-uniform hypergraphs, *Combin. Probab. Comput.* **15** (2006), no. 1–2, 229–251.

[RS06$^+$] V. Rödl and M. Schacht, Regular partitions of hypergraphs, submitted.

[RSST06$^+$] V. Rödl, M. Schacht, M. Siggers, and N. Tokushige, Integer and fractional packings of hypergraphs, submitted.

[RSTT06$^+$] V. Rödl, M. Schacht, E. Tengan, and N. Tokushige, Density theorems and extremal hypergraph problems, *Israel J. Math.*, to appear.

[RS06$^+$] V. Rödl and J. Skokan, Applications of the regularity lemma for uniform hypergraphs, *Random Structures Algorithms*, to appear.

[RS04] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, *Random Structures Algorithms* **25** (2004), no. 1, 1–42.

[Roth52] K. F. Roth, Sur quelques ensembles d'entiers, *C. R. Acad. Sci. Paris* **234** (1952), 388–390.

[Roth53] K. F. Roth, On certain sets of integers, *J. London Math. Soc.* **28** (1953), 104–109.

[RS78] I. Z. Ruzsa and E. Szemerédi, Triple systems with no six points carrying three triangles, *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976)*, Vol. II, North-Holland, Amsterdam, 1978, pp. 939–945.

[SS05] G. N. Sárközy and S. Selkow, On a Turán-type hypergraph problem of Brown, Erdős and T. Sós, *Discrete Math.* **297** (2005), no. 1-3, 190–195.

[Sol03] J. Solymosi, Note on a generalization of Roth's theorem, *Discrete and computational geometry, Algorithms Combin.*, vol. 25, Springer, Berlin, 2003, pp. 825–827.

[Sol04] J. Solymosi, A note on a question of Erdős and Graham, *Combin. Probab. Comput.* **13** (2004), no. 2, 263–267.

[Tao06$^+$a] T. Tao, The Gaussian primes contain arbitrarily shaped constellations, submitted.

[Tao06$^+$b] T. Tao, A variant of the hypergraph removal lemma, *J. Combinatorial Theory Ser. A*, to appear.

Part IV

Homomorphisms

Homomorphisms in Graph Property Testing

Noga Alon[1⋆] and Asaf Shapira[2⋆⋆]

[1] Schools of Mathematics and Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel
nogaa@tau.ac.il

[2] School of Computer Science, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel
asafico@tau.ac.il

Summary. Property-testers are fast randomized algorithms for distinguishing between graphs (and other combinatorial structures) satisfying a certain property, from those that are far from satisfying it. In many cases one can design property-testers whose running time is in fact *independent* of the size of the input. In this paper we survey some recent results on testing graph properties. A common thread in all the results surveyed is that they rely heavily on the simple yet useful notion of graph homomorphism. We mainly focus on the combinatorial aspects of property-testing.

AMS Subject Classification. 68R05, 68W20, 68W25.
Keywords. Graphs, Property Testing, Homomorphism, Regularity Lemma, Extremal Graph Theory.

1 Introduction

1.1 Property-testing Background

The meta problem in the area of property testing is the following: Design a randomized algorithm, which given a combinatorial structure S, can distinguish with high probability between the case that S satisfies some property $\mathcal{P}$ from the case that S is ϵ-far from satisfying $\mathcal{P}$. Here S is said to be ϵ-far from satisfying $\mathcal{P}$ if an ϵ-fraction of its representation should be modified in order to make S satisfy $\mathcal{P}$. The main goal is to design randomized algorithms, which look at a very small portion of the input, and using this information distinguish with high probability between the above two cases. Such algorithms are

⋆ Research supported in part by a USA Israeli BSF grant, by a grant from the Israel Science Foundation, and by the Hermann Minkowski Minerva Center for Geometry at Tel Aviv University.

⋆⋆ Research supported in part by a Charles Clore Foundation Fellowship and an IBM Ph.D Fellowship.

called *property testers* or simply *testers* for the property $\mathcal{P}$. Preferably, a tester should look at a portion of the input whose size is a function of ϵ only. Blum, Luby and Rubinfeld [BLR93] were the first to formulate a question of this type, and the general notion of property testing was first formulated by Rubinfeld and Sudan [RS96], who were motivated in studying various algebraic properties such as linearity of functions.

The main focus of the present survey is in testing properties of graphs. In this case a graph G is said to be ϵ-far from satisfying a property $\mathcal{P}$, if one needs to add/delete at least ϵn^2 edges to G in order to turn it into a graph satisfying $\mathcal{P}$. Here we assume that the tester can query an oracle, whether a pair of vertices, i and j, are adjacent in the input graph G. The study of the notion of testability for combinatorial structures, and mainly for labelled graphs, was introduced in the seminal paper of Goldreich, Goldwasser and Ron [GGR98]. In this paper it was shown that many natural graph properties such as k-colorability, having a large clique and having a large cut, have a tester, whose *query complexity* (that is, the number of oracle queries of type "does (i,j) belong to $E(G)$") can be upper bounded by a function of ϵ that is independent of the size of the input. In the present paper we will say that properties having such efficient testers, that is, testers whose query complexity is a function of ϵ only, are simply *testable*. Note, that if the query complexity of a tester can be upper bounded by a function of ϵ only, then so is its running time. Thus, if a property $\mathcal{P}$ is testable then there is a *constant* time randomized algorithm, for distinguishing between graphs satisfying $\mathcal{P}$ from those that are ϵ-far from satisfying it. In general, a property tester has a small probability of accepting graphs that are ϵ-far from satisfying the tested property, as well as a small probability of rejecting graphs satisfying the property. In this case the tester is said to have *two-sided* error. If the tester accepts graphs satisfying the property with probability 1, then the tester is said to have *one-sided* error.

We briefly mention that the model of graph property testing we study in this paper is usually referred to as the *dense graph model*. Other models of graph property testing have also been investigated, see [GR97] and [PR02]. For further reading and pointers on testing properties of graphs and other combinatorial structures the reader is referred to the surveys [Fish01], [Gol98], [Ron01] and their references.

1.2 Background on Graph Homomorphism

As the title of this survey alludes to, its main focus is applications of homomorphisms in graph property-testing. We start with defining homomorphism between graphs.

Definition 1.1 (Homomorphism). *A homomorphism from a graph F to a graph K is mapping $\varphi : V(F) \mapsto V(K)$, which maps edges to edges, namely $(v,u) \in E(F)$ implies $(\varphi(v), \varphi(u)) \in E(K)$.*

Throughout this survey, $F \mapsto K$ will denote that there is a homomorphism from F to K, and $F \not\mapsto K$ will denote that no such homomorphism exists. Practicing definitions, note that if $F \mapsto K$ then $\chi(F) \leq \chi(K)$, and that F has a homomorphism into the complete graph on k vertices if and only if F is k-colorable.

Though innocent looking, the notion of graph homomorphism comes in handy in many applications in graph theory and theoretical computer science. Note, that a homomorphisms between two graphs is not as informative as an isomorphism between them, and this lack of perfect information is useful in many situations.

It is useful to extend the standard notion of homomorphism to directed graphs as well. Thus, a homomorphism between two directed graphs is a function mapping vertices to vertices so that directed edges are mapped to directed edges (oriented in the same way). The following notion plays a crucial role in the results described in Section 8

Definition 1.2 (Core). *The core of a graph (or digraph) H, is the smallest (in terms of edges) subgraph K, of H, for which there is a homomorphisms from H to K.*

We refer the reader to [BH90] and [HNZ96] for more background and references on digraph homomorphisms, and to [HN92] for more information and references on cores of graphs. For additional reading on graph homomorphisms the reader is referred to the recent comprehensive book of Hell and Nešetřil [HN04] and its numerous references.

1.3 The Main Results Surveyed

In this survey we will mainly focus on the *combinatorial* aspects of property-testing rather than on its *algorithmic* aspects. To do so we will show that for the problems considered here, the property-testing tasks, which are algorithmic in nature, can be expressed in terms of (essentially equivalent) extremal graph-theoretic problems. Furthermore, all the properties surveyed here, will turn out to have testers whose query complexity is a function of ϵ only. This will make the property-testers we will consider extremely simple. In fact, all the algorithms for testing some property $\mathcal{P}$, will have the following meta-structure: Given a graph G and an error parameter ϵ: Sample a random set of vertices S, of size $f_{\mathcal{P}}(\epsilon)$ out of $V(G)$, and declare that G satisfies $\mathcal{P}$ if and only if the graph induced by G on S satisfies $\mathcal{P}$ (or is close to satisfying it in some cases). Of course, the main difficultly lies in showing that for a given property $\mathcal{P}$ there exists an $f_{\mathcal{P}}(\epsilon)$ for which such an algorithm distinguishes with high probability between graphs satisfying $\mathcal{P}$ and those that are ϵ-far from satisfying it.

This survey is organized as follows: In Section 2 we describe a result of [GT03] about the graph partition problems that are testable with one-sided error. Section 3 presents the regularity-lemma of Szemerédi [Sze78] and its applications, which are used in some of the other results surveyed in this paper. In Section 4 we discuss the main result of [AS05b], showing that any monotone graph property is testable. Section 5 is devoted to the main result of [ASS05], which gives approximation algorithms and hardness results for edge-deletion problems. In Section 7 we introduce the notion of colored-homomorphism, which was used in [AS05d] in order to prove that any hereditary graph property is testable, and in order to characterize the "natural" graph properties that are testable with one-sided error. In Section 6 we discuss the main result of [AS05c] about uniform and non-uniform property testing. In Section 8 we discuss the main results of [Alon02] and [AS04a] about testing H-freeness for a fixed directed or undirected graph H. Most of the sections contain some interesting open problems for future research.

2 Testing Partition Problems with One-sided Error

We start this section by introducing the notion of graph partitioning problems that was studied in [GGR98]: For some integer k, let S be a set of $k + k + \binom{k}{2}$ pairs of reals which we denote by (α_i, β_i) for every $1 \le i \le k$ and $(\gamma_{i,j}, \delta_{i,j})$ for every $1 \le i \le j \le k$. Then, a graph G on n vertices is said to satisfy the partition property $\mathcal{P}(S)$, if $V(G)$ can be partitioned into k sets of vertices $V_1, \ldots, V_k$ such that the following holds: (i) for every $1 \le i \le k$ we have $\alpha_i \le |V_i|/n \le \beta_i$. (ii) for every $1 \le i \le j \le k$ we have $\gamma_{i,j} \le |E(V_i, V_j)|/n^2 \le \delta_{i,j}$. In what follows, we will call k the *order* of S. One of the main results of [GGR98] states the following:

Theorem 2.1 ([GGR98]). *For any S of order k, $\mathcal{P}(S)$ can be tested with query complexity* $(1/\epsilon)^{c(k)}$.

At this point we should mention that a common theme in property testing (in the dense graph model, which is the focus of this survey) is that the testers themselves are quite simple, and the challenging part in designing them is showing that they indeed work. This is the main reason why we focus here on the combinatorial aspects of testing graph. In fact, for most problems the algorithm simply samples a set of vertices S, queries about all the pairs of vertices in S, and then accepts or rejects according to the graph spanned by S. For example, the tester for k-colorability, which is a partition problem as defined above, simply samples a set of vertices S of size $poly(1/\epsilon)$ and accepts if and only if the graph spanned by S is k-colorable. As another example, the tester for the property of having a cut of size at least $\frac{1}{8}n^2$, which is also a partition problem, samples a set of vertices S of size $poly(1/\epsilon)$ and accepts if and only if the graph spanned by S is has a cut of size at least $(\frac{1}{8} - \frac{\epsilon}{2})n^2$. The proof of Theorem 2.1 gives analogously simple algorithms for other partition problems. The testers designed in [GGR98] for general partition problems may have two-sided error. For some special cases, however, one can design one-sided error tester. One example is the k-colorability, which was shown to be testable with one-sided error in [GGR98]. One of the (three) results of [GT03], is a characterization of the partition problems, which can be tested with one-sided error and query complexity independent of the size of the input[1].

Theorem 2.2 ([GT03]). *A (non-trivial) partition problem $\mathcal{P}(S)$, is testable with one-sided error and query complexity bounded by a function of ϵ if and only if one of the following holds:*

1. *There is a graph H, such that satisfying $\mathcal{P}(S)$ is equivalent to having a homomorphism to H.*
2. *$\mathcal{P}(S)$ is equivalent to being a complete graph.*

It is easy to see that the property of being a complete graph is testable with one sided error and query-complexity $O(1/\epsilon)$. Given the techniques of [GGR98] and [AK02] it can be shown that for any fixed H, the property of having a homomorphism to H is testable with one sided error and query complexity $O(1/\epsilon^2)$. In fact, this result follows as a special case of the main result of [AS03]. The main difficulty in the proof of Theorem 2.2, thus lies in proving that any other partition problem cannot be tested with one-sided error. The details appear in [GT03]. It is worth noting that if one only considers "oblivious" testers, as defined in Section 7, then one can derive the negative side of Theorem 2.2 from Theorem 7.9. The reason is that it is not difficult to show that a partition problem defines a semi-hereditary graph property (see Section 7) if and only if it is either equivalent to the clique-property or is equivalent to having a homomorphism to a fixed graph H.

The most well-studied partition problem is probably k-colorability. For the case of $k = 2$ it is known that in order to test the property of being bipartite a sample of $\tilde{\Theta}(1/\epsilon)$ vertices is both sufficient and necessary, see [AK02]. However, for $k \geq 3$ it is only known that in order to test k-colorability one has to use a sample of size $\Omega(1/\epsilon)$ vertices, and one of size $O(1/\epsilon^2)$ suffices. It seems interesting to close this gap by resolving the following problem.

Open Problem 2.3. Prove tight bounds for the query complexity of k-colorability for $k \geq 3$.

3 Regularity Lemmas: Definitions, Statements and Applications

Several of the results discussed in this survey are obtained using the regularity-lemma of Szemerédi [Sze78]. In this section we give the necessary background

[1] The Characterization in [GT03] also considers degenerate cases, such as partition problems that are satisfied by only finitely many graphs. We focus here on the non-degenerate cases, which we call non-trivial.

on this powerful tool, which is needed in order to present these results. For a comprehensive survey on the regularity-lemma and its applications, the interested reader is referred to [KS96]. We start with some basic definitions. For every pair of nonempty disjoint vertex sets A and B of a graph G, we define $e(A, B)$ to be the number of edges of G between A and B. The *edge density* of the pair is defined by $d(A, B) = e(A, B)/|A||B|$.

Definition 3.1 (γ-regular pair). *A pair (A, B) is* γ-regular*, if for any two subsets $A' \subseteq A$ and $B' \subseteq B$, satisfying $|A'| \geq \gamma|A|$ and $|B'| \geq \gamma|B|$, the inequality $|d(A', B') - d(A, B)| \leq \gamma$ holds.*

Note, that a sufficiently large random bipartite graph, where each edge is chosen independently with probability d, is very likely to be a γ-regular pair with density roughly d, for any $\gamma > 0$. Thus, in some sense, the smaller γ is, the closer a γ-regular pair is to looking like a random bipartite graph. For this reason, the reader who is unfamiliar with the regularity lemma and its applications, should try and compare the statements given in this section to analogous statements about random graphs.

Let F be a graph on f vertices and K a graph on k vertices, and suppose $F \mapsto K$. Let G be a graph obtained by taking a copy of K, replacing every vertex with a sufficiently large independent set, and every edge with a random bipartite graph of edge density d. It is easy to show that with high probability G contains many copies of F. The following lemma shows that in order to infer that G contains many copies of F, it is enough to replace every edge with a "regular enough" pair. Intuitively, the larger f and k are, and the sparser the regular pairs are, the more regular we need each pair to be, because we need the graph to be "closer" to a random graph. This is formulated in Lemma 3.2 below. Several versions of this lemma were previously proved in papers using the regularity lemma. See, e.g., [KS96]. The reader should think of the mapping φ in the statement of the lemma as defining the homomorphism from F to the (implicit) graph K. In what follows, as well as throughout the paper, the notation $\gamma_{3.2}$ means the function/constant γ that was defined in Lemma/Claim/Thoerem 3.2.

Lemma 3.2 (The Embedding Lemma). *For every real $0 < \eta < 1$, and integers $k, f \geq 1$ there exist $\gamma = \gamma_{3.2}(\eta, k, f)$, $\delta = \delta_{3.2}(\eta, k, f)$ and $M = M_{3.2}(\eta, k, f)$ with the following property. Let F be any graph on f vertices, and let $U_1, \ldots, U_k$ be k pairwise disjoint sets of vertices, where $|U_1| = \ldots = |U_k| = m \geq M$. Suppose there is a mapping $\varphi : V(F) \mapsto \{1, \ldots, k\}$ such that the following holds: If (i, j) is an edge of F then $(U_{\varphi(i)}, U_{\varphi(j)})$ is γ-regular with density at least η. Then, the sets $U_1, \ldots, U_k$ span at least δm^f copies of F.*

A partition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of the vertex set of a graph is called an *equipartition* if $|V_i|$ and $|V_j|$ differ by no more than 1 for all $1 \leq i < j \leq k$ (so in particular each V_i has one of two possible sizes). The Regularity Lemma of Szemerédi can be formulated as follows.

Lemma 3.3 (Szemerédi's Regularity Lemma [Sze78]). *For every m and $\gamma > 0$ there exists a number $T = T_{3.3}(m,\gamma)$ with the following property: Any graph G on $n \geq T$ vertices, has an equipartition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of $V(G)$ with $m \leq k \leq T$, for which all pairs (V_i, V_j), but at most $\gamma\binom{k}{2}$ of them, are γ-regular.*

In most of the applications of Lemma 3.3 one removes from G the following three types of edges: (i) edges spanned by the sets $V_1,\ldots,V_k$ (ii) edges connecting pairs (V_i, V_j), whose density $d(V_i, V_j)$ is small, say, smaller than γ (iii) edges connecting pairs (V_i, V_j), which are not γ-regular. The main reason for disregarding these edges (at least in applications similar in nature to the ones considered here) is that we would want to apply Lemma 3.2 on some of the sets $V_1,\ldots,V_k$ and to use this lemma we need pairs that are both regular and dense. A simple, yet useful observation, is that if we apply Lemma 3.2 with $m > 1/\gamma$, then the total number of edges removed in (i),(ii) and (iii) above is smaller than $2\gamma n^2$. Thus we can restate the regularity lemma in the following more convenient way:

Lemma 3.4. *For every $\gamma > 0$ and $m > 1/\gamma$ there exists a number $T = T_{3.4}(m,\gamma)$ with the following property: from any graph G on $n > T$ vertices, one can remove at most $2\gamma n^2$ edges to obtain a k-partite graph, with partition classes $V_1,\ldots,V_k$, such that the following holds:*

1. *$m \leq k \leq T$.*
2. *Any pair (V_i, V_j) is γ-regular[2] and satisfies either $d(V_i, V_j) = 0$ or $d(V_i, V_j) \geq \gamma$.*

For a (possibly infinite) family of graphs $\mathcal{F}$, a graph G is said to be $\mathcal{F}$-free if it contains no $F \in \mathcal{F}$ as a (not necessarily induced) subgraph. We need the following simple lemma, which reduces the task of testing the property of being $\mathcal{F}$-free to a purely combinatorial problem.

Lemma 3.5. *Let $\mathcal{F}$ be a (possibly infinite) family of graphs, and suppose there are functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ such that the following holds for every $\epsilon > 0$: Every graph G on n vertices, which is ϵ-far from begin $\mathcal{F}$-free contains at least $\delta_{\mathcal{F}}(\epsilon)n^f$ copies of a graph $F \in \mathcal{F}$ of size $f \leq f_{\mathcal{F}}(\epsilon)$. Then, being $\mathcal{F}$-free is testable with one-sided error.*

Proof. Suppose $\mathcal{F}$ is a family for which the functions $f_{\mathcal{P}}(\epsilon)$ and $\delta_{\mathcal{P}}(\epsilon)$ exist. For any $\epsilon > 0$ put $f = f_{\mathcal{P}}(\epsilon)$ and $\delta = \delta_{\mathcal{P}}(\epsilon)$. Given a graph the tester picks a set of vertices S, of size $2f/\delta$ and accepts G if and only if the subgraph of G spanned by S is $\mathcal{F}$-free. Clearly, if G is $\mathcal{F}$-free the algorithm accepts G with probability 1. Assume now that G is ϵ-far from being $\mathcal{F}$-free. In this case a randomly chosen set of f vertices spans a copy of a graph $F \in \mathcal{F}$ with probability at least δ. Thus, a sample of size $2f/\delta$ spans a copy of F, with probability at least 2/3. □

[2] Note that if $d(V_i, V_j) = 0$ then (V_i, V_j) is trivially γ-regular.

A standard application of Lemmas 3.2 and 3.4 gives the following:

Lemma 3.6. *For any* **finite** *set of graphs $\mathcal{F}$, the property of being $\mathcal{F}$-free is testable.*

Proof (sketch). The strategy is quite simple: we would like to show that if a graph G is ϵ-far from being $\mathcal{F}$-free then there is a graph $F \in \mathcal{F}$ and t subsets of $V(G)$, denoted $V_1, \ldots, V_t$, such that Lemma 3.2 can be applied on these sets with respect to F. By Lemma 3.5 this will be sufficient for inferring that $\mathcal{F}$-freeness is testable.

To carry out the above strategy we first need to know how regular should the sets $V_1, \ldots, V_t$ be in order to let us use Lemma 3.2 on them. As $\mathcal{F}$ is finite, there is an upper bound, say f_{max}, on the size of the largest graph in $\mathcal{F}$. Thus, we would not need to apply Lemma 3.2 with $k > f_{max}$ or with $f > f_{max}$. Therefore, if the sets $V_1, \ldots, V_t$ are γ'-regular, where we set $\gamma' = \gamma_{3.2}(\epsilon, f_{max}, f_{max})$ (it will soon become clear why we use ϵ here), then we will be able to apply Lemma 3.2 on the sets $V_1, \ldots, V_t$ to find many copies of *any* $F \in \mathcal{F}$. But, to do so we still have to find the t sets $U_1, \ldots, U_t$ and argue why the densities between them correspond to the edge set of some graph $F \in \mathcal{F}$. To this end, we apply Lemma 3.4 with $\gamma = \frac{1}{2}\min(\epsilon, \gamma')$ and $m = 1/\gamma$. By Lemma 3.4, we can remove from G at most ϵn^2 edges and get a k-partite graph, denoted G', as in the statement of the lemma. As G is by assumption ϵ-far from being $\mathcal{F}$-free G' still spans a copy of some $F \in \mathcal{F}$. Suppose F is spanned by the sets $V_1, \ldots, V_t$ and recall that we must have $|V(F)| \leq f_{max}$ and $t \leq f_{max}$. Furthermore, by the definition of G' we have the following important property: if $x, y \in V(F)$ are connected in F, vertex x belongs to V_i and vertex y belongs to V_j, then (V_i, V_j) must be ϵ-regular and satisfy $d(V_i, V_j) \geq \epsilon$. As we chose $\gamma' = \gamma_{3.2}(\epsilon, f_{max}, f_{max})$, we can apply Lemma 3.2 on the sets $V_1, \ldots, V_t$. As each of these sets is of size at least $n/T_{3.4}(1/\gamma, \gamma)$, we conclude from Lemma 3.2 that $V_1, \ldots, V_t$ span at least

$$\delta_{3.2}(\epsilon, f_{max}, f_{max}) \left(\frac{n}{T_{3.4}(1/\gamma, \gamma)} \right)^{|V(F)|}$$

copies of F. As γ was defined in terms of γ', which is defined in terms ϵ we get from the above expression that G' spans some $\delta_{\mathcal{F}}(\epsilon) n^{|V(F)|}$. As G' is a subgraph of G, G contains at least this many copies of F. □

Let us try to apply the argument given in the above proof when $\mathcal{F}$ is an infinite set. The main problem is that there is no upper-bound on the size of the graphs in $\mathcal{F}$. Thus, we do not know a priori the size of the member of $\mathcal{F}$ that we will eventually find in the equipartition that Lemma 3.4 returns. After finding $F \in \mathcal{F}$ in an equipartition, we may find out that F is too large for Lemma 3.2 to be applied, because Lemma 3.4 was not used with a small enough γ. One may then try to find a new equipartition based on the size of F. However, that requires using a smaller γ, and thus the new equipartition may be larger (that is, contain more partition classes), and thus contain only larger members of $\mathcal{F}$. Hence, even the new γ is not good enough in order to apply Lemma 3.2. This leads to a circular definition of constants, which seems unbreakable. The main tool in the proof of Theorem 4.3 is Lemma 3.7 below, proved in [AFKS00] with a different motivation, which enables one to break this circular chain of definitions. This lemma can be considered a variant of the standard regularity lemma, where one can use a *function* that defines γ as a function of the size of the partition, rather then having to use a *fixed* γ as in Lemma 3.4.

Lemma 3.7 ([AFKS00]). *For every integer m and monotone non-increasing function $\mathcal{E}(r) : \mathbb{N} \mapsto (0,1)$ there is an integer $S = S_{3.7}(m, \mathcal{E})$ satisfying the following: For any graph G on $n \geq S$ vertices, there exist an equipartition $\mathcal{A} = \{V_i \mid 1 \leq i \leq k\}$ of $V(G)$ and an induced subgraph U of G, with an equipartition $\mathcal{B} = \{U_i \mid 1 \leq i \leq k\}$ of the vertices of U, that satisfy:*

1. *$m \leq k \leq S$.*
2. *$U_i \subseteq V_i$ for all $i \geq 1$, and $|U_i| \geq n/S$.*
3. *In the equipartition $\mathcal{B}$,* **all** *pairs are $\mathcal{E}(k)$-regular.*
4. *All but at most $\mathcal{E}(0)\binom{k}{2}$ of the pairs $1 \leq i < j \leq k$ are such that $|d(V_i, V_j) - d(U_i, U_j)| < \mathcal{E}(0)$.*

The dependency of the function $T_{3.3}(m, \gamma)$ on γ is a tower of exponents of height polynomial in $1/\gamma$ (see the proof in [KS96]). As one can infer from the details in [AFKS00], this implies that even for moderate functions $\mathcal{E}$ the integer $S(m, \mathcal{E})$ grows as fast as a tower of towers of exponents (see the proof in [AFKS00]).

4 Testing Monotone Graph Properties

The problem of characterizing the testable graph properties is widely considered to be the most important open problem in the area of property testing. A natural step in this direction is in showing that large and natural families of graph properties are testable. In this section we consider the family of monotone graph properties. A graph property is *monotone* if it is closed under removal of vertices and edges, or equivalently if it is closed under taking (not necessarily induced) subgraphs. Various monotone graph properties were extensively studied in graph theory. As examples of monotone properties one can consider the property of having a homomorphism to a fixed graph H (which includes as a special case the property of being k-colorable, see Definition 1.1), and the property of not containing a (not necessarily induced) copy of some fixed graph H. Another monotone property is being $(k, \mathcal{H})$-*Ramsey*: For a (possibly infinite) family of graphs $\mathcal{H}$, a graph is said to be $(k, \mathcal{H})$-Ramsey if one can color its edges using k colors, such that no color class contains a copy of a graph $H \in \mathcal{H}$. This property is the main focus of Ramsey-Theory, see [GRS90] and its references. As another example, one can consider the property of being $(k, \mathcal{H}, f)$-Multicolorable; For a (possibly infinite) family of graphs $\mathcal{H}$ and a function f from $\mathcal{H}$ to the positive integers, a graph is said to be $(k, \mathcal{H}, f)$-Multicolorable if one can color its edges using k colors, such that every copy of a graph $H \in \mathcal{H}$ receives at least $f(H)$ colors. See [EG97], [EM00] and their references for a discussion of some special cases. Another set of well studied monotone properties are those defined by having a bounded *fractional chromatic number*, bounded *vector chromatic number* (see [KMS98]) or bounded *Lovász theta function* (see [Lov79]). The abstract family of monotone graph properties has also been extensively studied in graph theory. See [FK96], [BBW06$^+$], [AS00] and their references. The main focus of this section is in presenting the main ideas behind the proof of the following result:

Theorem 4.1 ([AS05b]). *Every monotone graph property is testable with one-sided error.*

It is important to note that the main result of [AS05b] is slightly weaker than the statement given in Theorem 4.1 as the monotone property has to satisfy some technical conditions (which cannot be avoided). Roughly speaking, the main problem is that for some properties it may be impossible (that is, non-recursive) given ϵ, to compute the number of queries the tester should perform in order to test the monotone property. See Section 6 for the precise statement and discussion of this issue. In this section we will overlook this issue and focus on the combinatorial part of the problem. Another important note is that in [GT03], Goldreich and Trevisan define a monotone graph property to be one that is closed under removal of edges, and not necessarily under removal of vertices. They show that there are such properties that are not testable even with two sided error. In fact, their result is stronger as the property they define belongs to NP and requires query complexity $\Omega(n^2)$. This means that Theorem 4.1 cannot be extended, in a strong sense, to properties that are only closed under removal of edges. As we show in Section 7, Theorem 4.1 can be extended to properties, which are closed only under removal of vertices, namely, hereditary properties.

We next introduce a convenient way of handling a monotone graph property.

Definition 4.2 (Forbidden Subgraphs). *For a monotone graph property $\mathcal{P}$, define $\mathcal{F} = \mathcal{F}_{\mathcal{P}}$ to be the set of graphs which are minimal with respect to not satisfying property $\mathcal{P}$. In other words, a graph F belongs to $\mathcal{F}$ if it does not satisfy $\mathcal{P}$, but any graph obtained from F by removing an edge or a vertex, satisfies $\mathcal{P}$.*

As an example of a family of forbidden subgraphs, observe that if $\mathcal{P}$ is the property of being 2-colorable, then $\mathcal{F}_{\mathcal{P}}$ is the set of all odd-cycles. Clearly, a graph satisfies $\mathcal{P}$ if and only it contains no member of $\mathcal{F}_{\mathcal{P}}$ as a (not necessarily induced) subgraph. Recall, that as was defined in Section 3, a graph is $\mathcal{F}$-free if it contains no (not necessarily induced) subgraph $F \in \mathcal{F}$. Clearly, for any family $\mathcal{F}$, being $\mathcal{F}$-free is a monotone property. Thus, the monotone properties are precisely the graph properties of being $\mathcal{F}$-free for some family $\mathcal{F}$. In order to simplify the notation of this section, it will be simpler to talk about properties of type $\mathcal{F}$-free rather than monotone properties. The reader should keep in mind that we allow $\mathcal{F}$ to be an infinite set. By Lemma 3.5 and the above discussion we can prove Theorem 4.1 by proving the following:

Theorem 4.3. *For any (possibly infinite) family of graphs $\mathcal{F}$ there are functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ satisfying the conditions of Lemma 3.5.*

We remind the reader that as was shown in Section 3, the standard regularity-lemma can be applied in order to prove Theorem 4.3 for any *finite* family $\mathcal{F}$. In order to prove Theorem 4.3 for infinite families of graphs we apply the strengthening of the regularity-lemma, which is stated in Lemma 3.7. A sketch of the proof of Theorem 4.3 is given in Subsection 4.1 below. Before getting into the details we discuss some interesting applications of Theorem 4.3. First, observe that an immediate corollary of this theorem is the following:

Corollary 4.4 ([AS05b]). *For every monotone graph property $\mathcal{P}$, there is a function $W_{\mathcal{P}}(\epsilon)$ with the following property: If G is ϵ-far from satisfying $\mathcal{P}$, then G contains a subgraph of size at most $W_{\mathcal{P}}(\epsilon)$, which does not satisfy $\mathcal{P}$.*

The above corollary significantly extends a result of Rödl and Duke [RD85], conjectured by Erdős, which asserts that the above statement holds for the k-colorability property. Corollary 4.4 extends this result to any monotone property and in particular to the monotone properties discussed at the beginning of this section.

As the details of the proof of Theorem 4.3 reveal the functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ depend on the family of graphs $\mathcal{F}$. This means, that for any monotone graph property $\mathcal{P}$ the upper bounds for testing $\mathcal{P}$, which Theorem 4.1 guarantees depend on the specific property being tested. A natural question one may ask, is if the dependency on the specific property being tested can be removed. One can rule out this possibility by proving the following.

Theorem 4.5 ([AS05b]). *For any function $Q : (0,1) \mapsto \mathbb{N}$, there is a monotone graph property $\mathcal{P}$, which cannot be tested with one-sided error and query complexity $o(Q(\epsilon))$.*

Prior to [AS05b], the best lower bound proved for testing a testable graph property with one-sided error was obtained in [Alon02], where it is shown that for every non-bipartite graph H, the query complexity of testing whether a graph does not contain a copy of H is at least $(1/\epsilon)^{\Omega(\log 1/\epsilon)}$ (see Section 8). The fact that for every H this property is testable with one-sided error, follows from [ADLRY94] and [AFKS00], and also as a special case of Theorem 4.1.

As by Theorem 4.1 every monotone graph property is testable with one-sided error, Theorem 4.5 establishes that the one-sided error query complexity of testing testable graph properties, even those that are testable with one-sided error, may be *arbitrarily large*. We finally note that the proof of Theorem 4.5 implies a similar statement with respect to Corollary 4.4.

Another application of Theorem 4.3 can be considered a compactness-type result in property testing. Suppose $\mathcal{P}_1, \ldots, \mathcal{P}_k$ are k graph properties that are closed under removal of edges. It is clear that if a graph G is ϵ-far from satisfying these k properties then it is at least ϵ/k-far from satisfying at least one of them. However, it is not clear that there is a fixed $\delta > 0$ such that even if $k \to \infty$, G must be δ-far from satisfying one of these properties. By using Theorem 4.3 one can prove that if these properties are monotone then such an δ exists. It can also be shown that in general, no such δ exists. This is stated in the following theorem, see [AS05b].

Theorem 4.6 ([AS05b]). *For any (possibly infinite) set of monotone graph properties $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots\}$, there is a function $\delta_{\mathcal{P}} : (0,1) \mapsto (0,1)$ with the following property: If a graph G is ϵ-far from satisfying all the properties of $\mathcal{P}$, then for some i, the graph G is $\delta_{\mathcal{P}}(\epsilon)$-far from satisfying $\mathcal{P}_i$. Furthermore, there are properties $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots\}$, which are closed under removal of edges for which no such $\delta_{\mathcal{P}}$ exists.*

We turn to some interesting open problems, which the above results suggest. As was mentioned above, a result of Goldreich and Trevisan [GT03] rules out the possibility of extending Theorem 4.1 to graph properties that are only closed under removal of edges. It seems interesting to bridge the gap between their result and Theorem 4.1 by resolving the following problem.

Open Problem 4.7. Characterize the testable graph properties that are closed under edge removal.

It also seems interesting to see if the new powerful hypergraph versions of the regularity lemma (see [Gow06$^+$], [NRS06$^+$] and [RS04]) can be used to obtain hypergraph versions of Lemma 3.7, and if in that case, one can obtain hypergraph versions of Theorem 4.1 and Corollary 4.4.

Open Problem 4.8. Prove hypergraph versions of Theorem 4.1 and Corollary 4.4.

It may also be interesting to extend Theorem 4.5 by showing the following:

Open Problem 4.9. Prove that for any function $Q : (0,1) \mapsto \mathbb{N}$, there is a monotone graph property $\mathcal{P}$, which cannot be tested (either with one-sided or two-sided error) with query complexity $o(Q(\epsilon))$.

4.1 Sketch of the Proof of Theorem 4.3

In this subsection we give an overview of the proof of Theorem 4.3. The reader is referred to [AS05b] for the full proof. Though this result may seem to have nothing to do with homomorphisms, the key idea in the proof is in using a certain graph functional that involves graph homomorphisms.

We first need to introduce another standard notion frequently used in applications of the regularity-lemma. For an equipartition of a graph G, let the *regularity graph* of G, denoted $R = R(G)$, be the following graph: We first use Lemma 3.4 in order to obtain the equipartition satisfying the assertions of the lemma. Let k be the size of the equipartition. Then, R is a graph on k vertices, where vertices i and j are connected if and only if (V_i, V_j) is γ-regular and $d(V_i, V_j) \geq \gamma$. In some sense, the regularity graph is an approximation of the original graph, up to γn^2 modifications. One of the main (implicit) implications of the regularity lemma is the following: Suppose we consider two graphs to be *similar* if their regularity graphs are isomorphic. It thus follows from Lemma 3.4 that for every $\gamma > 0$, the number of graphs that are pairwise non-similar is bounded by a function of γ, which is roughly $2^{\binom{T}{2}}$, where $T = T_{3.4}(1/\gamma, \gamma)$. In other words, up to γn^2 modifications, all the graphs can be approximated using a set of equipartitions of size bounded by a function of γ only. One (easy) application of this observation is that there are $2^{(\frac{1}{4}+o(1))n^2}$ triangle-free graphs on n vertices. The reader is referred to [EFR86] where this interpretation of the regularity lemma is further (implicitly) used. This leads us to the key definitions of the proof of Theorem 4.3. The reader should think of the graphs R considered below as the set of regularity graphs discussed above, and the parameter r as representing the size of R.

Definition 4.10 (The family $\mathcal{F}_r$). *For any (possibly infinite) family of graphs $\mathcal{F}$, and any integer r let $\mathcal{F}_r$ be the following set of graphs: A graph R belongs to $\mathcal{F}_r$ if it has at most r vertices and there is at least one $F \in \mathcal{F}$ such that $F \mapsto R$.*

Practicing definitions, observe that if $\mathcal{F}$ is the family of odd cycles, then $\mathcal{F}_r$ is precisely the family of non-bipartite graphs of size at most r. In the proof of Theorem 4.3, the set $\mathcal{F}_r$ will represent a subset of the regularity graphs of size at most r. Namely, those R for which there is at least one $F \in \mathcal{F}$ such that $F \mapsto R$. As r will be a function of ϵ only, and thus finite, we can take the maximum over all the graphs $R \in \mathcal{F}_r$, of the size of the smallest $F \in \mathcal{F}$ such that $F \mapsto R$. We thus define

Definition 4.11 (The function $\Psi_{\mathcal{F}}$). *For a family of graphs $\mathcal{F}$ and an integer r for which $\mathcal{F}_r \neq \emptyset$, let*

$$\Psi_{\mathcal{F}}(r) = \max_{R \in \mathcal{F}_r} \min_{\{F \in \mathcal{F} : F \mapsto R\}} |V(F)|. \tag{1}$$

Define $\Psi_{\mathcal{F}}(r) = 0$ if $\mathcal{F}_r = \emptyset$. Therefore, $\Psi_{\mathcal{F}}(r)$ is monotone non-decreasing in r.

Practicing definitions again, note that if $\mathcal{F}$ is the family of odd cycles, then $\Psi_{\mathcal{F}}(r) = r$ when r is odd, and $\Psi_{\mathcal{F}}(r) = r - 1$ when r is even. The "right" way to think of the function $\Psi_{\mathcal{F}}$ is the following: Let R be a graph of size at most r and suppose we are guaranteed that there is a graph $F' \in \mathcal{F}$ such that $F' \mapsto R$ (thus $R \in \mathcal{F}_r$). Then, by this information only and *without* having to know the structure of R itself, the definition of $\Psi_{\mathcal{F}}$ implies that there is a graph $F \in \mathcal{F}$ of size at most $\Psi_{\mathcal{F}}(r)$, such that $F \mapsto R$.

The function $\Psi_{\mathcal{F}}$ has a critical role in the proof of Theorem 4.3. The first usage of this function is that as by Lemma 3.7 we can upper bound the size of the regularity graph R, we can also upper bound the size of the smallest graph $F \in \mathcal{F}$ for which $F \mapsto R$. A second important property of $\Psi_{\mathcal{F}}$ is discussed in Section 6. As we have mentioned in the previous section, the main difficulty that prevents one from proving Theorem 4.3 using Lemma 3.2 is that one does not know a priori the size of the graph that one may expect to find in the equipartition. This leads us to define the following function where $0 < \epsilon < 1$ is an arbitrary real.

$$\mathcal{E}(r) = \begin{cases} \epsilon, & r = 0 \\ \gamma_{3.2}(\epsilon,\ r,\ \Psi_{\mathcal{F}}(r)), & r \geq 1 \end{cases} \tag{2}$$

In simple words, given $r > 0$, which will represent the size of the equipartition and thus also the size of the regularity graph which it defines, $\mathcal{E}(r)$ returns "how regular" this equipartition should be in order to allow one to find many copies of the *largest* graph one may possibly have to work with. Note, that we obtain the upper bound on the size of this largest possible graph, by invoking $\Psi_{\mathcal{F}}(r)$. As for different families of graphs $\mathcal{F}$, the function $\Psi_{\mathcal{F}}(r)$ may behave differently, $\mathcal{E}(r)$ may also behave differently for different families $\mathcal{F}$, as it is defined in terms of $\Psi_{\mathcal{F}}(r)$. However, and this is one of the key points of the proof, as we are fixing the family of graphs $\mathcal{F}$, the function $\mathcal{E}(r)$ depends only on r and implicitly on ϵ.

Given the above definitions we apply Lemma 3.7 with $\mathcal{E}(r)$ in order to obtain an equipartition of G. We then throw away edges that reside inside the sets V_i and between (V_i, V_j) whose edge density is either small or differs significantly from that of (U_i, U_j) (similar, but not identical, to what we did in Lemma 3.4). We then argue that we thus throw away less than ϵn^2 edges. As G is by assumption ϵ-far from not containing a member of $\mathcal{F}$, the new graph still contains a copy of $F \in \mathcal{F}$. By the definition of the new graph, it thus means that there is a (natural) homomorphism from F to the regularity graph of G. We then arrive at the main step of the proof, where we use the key property of Lemma 3.7, item (3), and the definition of $\mathcal{E}(r)$ to get that the sets U_i are regular enough to let us use Lemma 3.2 on them and to infer that they span many copies of F.

5 Approximation Algorithms for Edge-deletion Problems

The main topic of this section is graph modification problems, namely problems of the type: "given a graph G, find the smallest number of modifications that are needed in order to turn G into a graph satisfying property $\mathcal{P}$". The main two types of such problems are the following; In *node modification* problems, one tries to find the smallest set of vertices, whose removal turns G into a graph satisfying $\mathcal{P}$, while in *edge modification* problems, one tries to find the smallest number of edge deletions/additions, which turn G into a graph satisfying $\mathcal{P}$. In this section we will focus on edge-modification problems. Note, that when trying to turn a graph into one satisfying a monotone property we will only use edge deletions. Therefore, in these cases the problem is sometimes called *edge-deletion* problem. Before continuing, we need to introduce some notations. For a graph property $\mathcal{P}$, let $\mathcal{P}_n$ denote the set of graphs on n vertices, which satisfy $\mathcal{P}$. Given two graphs on n vertices, G and G', we denote by $\Delta(G, G')$ the edit distance between G and G', namely the smallest number of edge additions and/or deletions that are needed in order to turn G into G'. For a given property $\mathcal{P}$, we want to denote how far a graph G is from satisfying $\mathcal{P}$. For notational reasons it will be more convenient to normalize this measure so that it is always in the interval $[0, 1]$ (actually $[0, \frac{1}{2}]$). We thus define

Definition 5.1 ($E_{\mathcal{P}}(G)$). *For a graph property $\mathcal{P}$ and a graph G on n vertices, let*

$$E_{\mathcal{P}}(G) = \min_{G' \in \mathcal{P}_n} \frac{\Delta(G, G')}{n^2} .$$

In words, $E_{\mathcal{P}}(G)$ is the minimum edit distance of G to a graph satisfying $\mathcal{P}$ after normalizing it by a factor of n^2. The main result discussed in the previous section, Theorem 4.1, can be rephrased as follows: For any monotone property $\mathcal{P}$, one can distinguish between graphs satisfying $E_{\mathcal{P}}(G) = 0$, from those satisfying $E_{\mathcal{P}}(G) > \epsilon$. In this section we consider the natural extension of this problem, which asks for actually computing $E_{\mathcal{P}}(G)$, or to at least approximating it.

Graph modification problems have been studied extensively. Already in 1979, Garey and Johnson [GJ79] mentioned 18 types of vertex and edge modification problems. Graph modification problems were extensively studied as these problems have applications in several fields, including Molecular Biology and Numerical Algebra. In these applications a graph is used to model experimental data, where edge modifications correspond to correcting errors in the data: Adding an edge means correcting a false negative, while deleting an edge means correcting a false positive. Computing $E_{\mathcal{P}}(G)$ for appropriately defined properties $\mathcal{P}$ have important applications in physical mapping of DNA (see [CMNR97], [GGKS95] and [GKS94]). Computing $E_{\mathcal{P}}(G)$ for other properties arises when optimizing the running time of performing Gaussian elimination on a sparse symmetric positive-definite matrix (see [Rose72]). Other modification problems arise as subroutines for heuristic algorithms for computing the largest clique in a graph (see [Xue94]). Some edge modification problems also arise naturally in optimization of circuit design [EC88]. We briefly mention that there are also many results about *vertex* modification problems, notably that of Lewis and Yannakakis [LY80], who proved that for any non-trivial hereditary property $\mathcal{P}$, it is NP-hard to compute the smallest number of vertex deletions, which turn a graph into one satisfying $\mathcal{P}$.

As in Section 4 it will be simpler to consider a monotone property via its family of forbidden subgraphs $\mathcal{F}$ (see Definition 4.2). Therefore, $E_{\mathcal{F}}(G)$ will denote $E_{\mathcal{P}}(G)$, where $\mathcal{P}$ is the property of being $\mathcal{F}$-free. The main idea behind the algorithm for approximating $E_{\mathcal{F}}(G)$ is very simple: Given G and ϵ we would like to construct a small weighted (complete) graph W, of size depending on ϵ only, such that $E_{\mathcal{F}}(G)$ is close to some natural function of W. Surprisingly, again, this function is related to homomorphisms. We need the following definitions:

Definition 5.2 ($\mathcal{F}$-homomorphism-free). *For a family of graphs $\mathcal{F}$, a graph W is called $\mathcal{F}$-homomorphism-free if $F \not\mapsto W$ for any $F \in \mathcal{F}$.*

Definition 5.3 ($\mathcal{H}_{\mathcal{F}}(W)$). *For a family of graphs $\mathcal{F}$ and a weighted complete graph W of size k, let $\mathcal{H}'_{\mathcal{F}}(W)$ denote the minimum total weight of a set of edges, whose removal from W turns it into an $\mathcal{F}$-homomorphism-free graph. Let $\mathcal{H}_{\mathcal{F}}(W) = \frac{1}{k^2}\mathcal{H}'_{\mathcal{F}}(W)$.*

Note, that in Definition 5.2 the graph W is an unweighed not necessarily complete graph. The following is one of the main (implicit) results of [ASS05]:

Theorem 5.4 ([ASS05]). *For any family of graphs $\mathcal{F}$ and any $\epsilon > 0$ there is a deterministic $O(|V|+|E|)$ time algorithm with the following property: Given any graph $G = (V, E)$, the algorithm constructs a complete weighted graph W, of size $c = c(\epsilon, \mathcal{F})$ such that*

$$|E_{\mathcal{F}}(G) - \mathcal{H}_{\mathcal{F}}(W)| \leq \epsilon \ .$$

After obtaining the graph W using the above theorem, we can use exhaustive search in order to precisely compute $\mathcal{H}_{\mathcal{F}}(W)$. Note, that the time needed to precisely compute $\mathcal{H}_{\mathcal{F}}(W)$ is only a function of ϵ, because $|V(W)|$ is independent of the size of the input graph. This implies the following:

Corollary 5.5 ([ASS05]). *For any fixed $\epsilon > 0$ and any monotone property $\mathcal{P}$ there is a* **deterministic** *algorithm that given a graph $G = (V, E)$ computes in time $O(|V| + |E|)$ a real E satisfying $|E - E_{\mathcal{P}}(G)| \leq \epsilon$.*

Observe, that the running time of the algorithm on an n vertex dense graph is of type $f(\epsilon)n^2$. We note that Corollary 5.5 was not known for many monotone properties. In particular, such an approximation algorithm was not even known for the property of being triangle-free (and more generally for the property of being H-free for any non-bipartite H).

Theorem 5.4 is proved in [ASS05] via a novel structural graph theoretic technique, whose proof is not included here. We just mention that it uses some of the ideas used to prove Theorem 4.1 along with several additional ideas. It is also the first result to make a non-trivial application of the algorithmic version of the regularity lemma stated in Lemma 3.7. An additional interesting application of the structural result of [ASS05] is the following concentration-type result regarding the edit distance of small subgraphs of a graph.

Theorem 5.6 ([ASS05]). *For every $\epsilon > 0$ and any monotone property $\mathcal{P}$ there is $r = r(\epsilon, \mathcal{P})$ with the following property: Let G be any graph, and suppose we randomly pick a subset R, of r vertices from $V(G)$. Denote by G' the graph induced by G on R. Then,*

$$Prob[\ |E_{\mathcal{P}}(G') - E_{\mathcal{P}}(G)| > \epsilon] < \epsilon\ .$$

Again, once we have G', whose size depends only on ϵ, we can use exhaustive search in order to precisely compute $E_{\mathcal{P}}(G')$. Therefore, an immediate implication of the above theorem is the following extension of Theorem 4.1:

Corollary 5.7 ([ASS05]). *For every fixed $\epsilon > 0$ and any monotone property $\mathcal{P}$ there is a* **randomized** *algorithm, which given a graph G computes in time $O(1)$ a real E satisfying $|E - E_{\mathcal{P}}(G)| \le \epsilon$ with high probability.*

In standard Property-Testing one wants to distinguish between the graphs G that satisfy a certain graph property $\mathcal{P}$, or equivalently those G for which $E_{\mathcal{P}}(G) = 0$, from those that satisfy $E_{\mathcal{P}}(G) > \epsilon$. Parnas, Ron and Rubinfeld [PRR04] introduced the notion of Tolerant Property-Testing, where one wants to distinguish between the graphs G that satisfy $E_{\mathcal{P}}(G) < \delta$ from those that satisfy $E_{\mathcal{P}}(G) > \epsilon$, where $0 \le \delta < \epsilon \le 1$ are some constants. Recently, there have been several results in this line of work. Specifically, Fischer and Newman [FN05] have shown that if a graph property is testable with a number of queries depending on ϵ only, then it is also tolerantly testable for any $0 \le \delta < \epsilon \le 1$ and with query complexity depending on $|\epsilon - \delta|$. Combining this with Theorem 4.1 implies that any monotone property is tolerantly testable for any $0 \le \delta < \epsilon \le 1$ and with query complexity depending on $|\epsilon - \delta|$. Note, that Corollary 5.7 implicitly states the same. In fact, the algorithm implied by Corollary 5.7 is the "natural" one, where one picks a random subset of vertices S, and approximates $E_{\mathcal{P}}(G)$ by computing $E_{\mathcal{P}}$ on the graph induced by S. The algorithm of [FN05] is far more complicated. Furthermore, due to the nature of our algorithm if the input graph satisfies a monotone property $\mathcal{P}$, namely if $E_{\mathcal{P}}(G) = 0$, we will always detect that this is the case. The algorithm of [FN05] may declare that $E_{\mathcal{P}}(G) > 0$ even if $E_{\mathcal{P}}(G) = 0$.

Theorem 5.4 implies that it is possible to efficiently approximate the distance of an n vertex graph from any monotone graph property $\mathcal{P}$, to within an error of ϵn^2 for any $\epsilon > 0$. A natural question one can ask is for which monotone properties it is possible to improve the additive error to $n^{2-\delta}$ for some fixed $\delta > 0$. In the terminology of Definition 5.1, this means to approximate $E_{\mathcal{P}}$ to within an additive error of $n^{-\delta}$ for some $\delta > 0$ in polynomial time. In [ASS05], an essentially precise characterization of the monotone graph properties for which such a $\delta > 0$ exists, was given.

Theorem 5.8 ([ASS05]). *Let $\mathcal{P}$ be a monotone graph property. Then,*

1. *If there is a bipartite graph that does not satisfy $\mathcal{P}$, then there is a fixed $\delta > 0$ for which it is possible to approximate $E_{\mathcal{P}}$ to within an additive error of $n^{-\delta}$ in polynomial time.*
2. *On the other hand, if all bipartite graphs satisfy $\mathcal{P}$, then for any fixed $\delta > 0$ it is NP-hard to approximate $E_{\mathcal{P}}$ to within an additive error of $n^{-\delta}$.*

While the first part of the above theorem follows easily from the known results about the Turán numbers of bipartite graphs (see, e.g., [West01]), the proof of the second item involves various combinatorial tools including Szemerédi's Regularity Lemma, a new result in Extremal Graph Theory that extends the main result of [BSTT06+] and [BKS06+], and the basic approach of [Alon06+]. The proof of Theorem 5.8 is not included here, and the reader is referred to [ASS05] for the details.

For a fixed graph H, let $\mathcal{P}_H$ denote the property of being H-free. Note, that the above theorem implies that for any non-bipartite H, computing $E_{\mathcal{P}_H}$ is NP-hard. Observe, that this does not hold for the family of bipartite graphs H, as for some bipartite graphs such as the single edge, computing $E_{\mathcal{P}_H}$ can be done in polynomial time. It may thus be interesting to consider the following problem:

Open Problem 5.9. Characterize the bipartite graphs H, for which computing $E_{\mathcal{P}_H}$ is NP-hard.

6 Uniform versus Non-uniform Property Testing

One of the fundamental goals of complexity theory is in understanding the relations between various models of computation. In particular, one would like to know if two models are equivalent or if there are problems, which can be solved in one model but not in the other. Regretfully, in many interesting cases, though it seems intuitively obvious that two models of computation are not equivalent, the current techniques are far from enabling one to formally prove that. In this section we discuss two models of property-testing, which were introduced in [AS05c]. Surprisingly, in this case, though it seems at first that these models are equivalent, the main result of [AS05c] shows that it is possible to formally prove that they are in fact distinct (without making any hardness-type assumptions). The proof of this result relies heavily on the notion of graph homomorphism.

In defining a tester in Section 1 we did not mention, whether the error parameter ϵ is given as part of the input of the tester, or whether the tester is designed to distinguish between graphs that satisfy $\mathcal{P}$ from those that are ϵ-far from satisfying it, when ϵ is a known fixed constant. Prior to [AS05c] the literature about property testing was not clear about this issue as in some papers ϵ was assumed to be a part of the input while in others it is not. In [AS05c], the following two notions of property-testing were introduced:

Definition 6.1 (Uniformly testable). *A graph property $\mathcal{P}$ is* uniformly testable *if there is a tester that given ϵ, can distinguish with probability* 2/3 *between graphs that satisfy $\mathcal{P}$ and those that are ϵ-far from satisfying it, and whose query complexity is a function of ϵ only*[3]*. $\mathcal{P}$ is uniformly testable with one-sided error if the above tester accepts with probability 1 any graph satisfying $\mathcal{P}$.*

Definition 6.2 (Non-uniformly testable). *A graph property $\mathcal{P}$ is* non-uniformly testable *if for every fixed ϵ there is a tester that can distinguish with probability* 2/3 *between graphs that satisfy $\mathcal{P}$ and those that are ϵ-far from satisfying it, and whose query complexity is a constant (that is, a function of ϵ only). $\mathcal{P}$ is non-uniformly testable with one-sided error if each of the above testers accepts with probability 1 any graph satisfying $\mathcal{P}$.*

Note, that in the above two definitions the testers are not given the size of the input. This type of oblivious testers were defined and studied in [AS05d], where it was observed that all natural properties can be tested by testers that do not accept the size of the input. Therefore, for natural properties it seems that these testers are as powerful as testers accepting the size of the graph as part of the input. See Section 7 for more details on these slightly restricted testers.

It may seem, at least at first glance, that uniformly and non-uniformly property-testing are identical notions. The reader should note that the difference between being uniformly testable and non-uniformly testable, is not as sharp as, say, the difference between P and $P/Poly$ (see [Pap94]). The reason is that in $P/Poly$ the non-uniformity is with respect to the *inputs*, while in our case the non-uniformity is over the *error parameter*. In particular, a non-uniform tester for some fixed ϵ should be able to handle *any* input graph. The main result of [AS05c] shows that the above two notions of property testing are in fact distinct. Moreover, these notions are shown to be distinct while confining to graph properties, which are natural with respect to both their combinatorial structure and their computational difficulty. This is stated in the following result:

Theorem 6.3 ([AS05c]). *There is a graph property $\mathcal{P}$ with the following properties:*

[3] Note, that here and in the following definition we require that the query complexity of the tester will not only be *upper bounded* by a function of ϵ (as defined in Section 1), but actually be a function of ϵ only.

1. $\mathcal{P}$ can be non-uniformly tested with one-sided error.
2. $\mathcal{P}$ cannot be uniformly tested, even with two-sided error.

Moreover, satisfying $\mathcal{P}$ belongs to coNP and can be expressed in terms of forbidden subgraphs.

The property $\mathcal{P}$, which is constructed in [AS05c] in order to prove Theorem 6.3, is simply the property of being $\mathcal{F}$-free for some carefully defined family of graphs $\mathcal{F}$. We next give an overview of the proof of Theorem 6.3. The proof of this theorem heavily relies on the main result about testing monotone graph properties discussed in Section 4. As we have mentioned in Section 4 the actual result obtained in [AS05b] is slightly weaker than what was stated in Theorem 4.1. We are now ready to state the precise result of [AS05b]. In the statement we refer to the function $\Psi_{\mathcal{F}}$ defined in Definition 4.11. We also call a function *recursive* if there is an algorithm for computing it in finite time (see [Pap94]).

Theorem 6.4 ([AS05b]). *For every (possibly infinite) family of graphs $\mathcal{F}$, the property of being $\mathcal{F}$-free is non-uniformly testable with one-sided error. Moreover, if $\Psi_{\mathcal{F}}$ is recursive then being $\mathcal{F}$-free is also uniformly testable with one-sided error.*

The proof of Theorem 6.3 consists of two steps. The first establishes the somewhat surprising fact, that the property of $\Psi_{\mathcal{F}}(k)$ being recursive is not only sufficient for inferring that being $\mathcal{F}$-free is uniformly testable (as is stated in Theorem 6.4), but this condition is also necessary. This is formulated in the following Theorem.

Theorem 6.5 ([AS05b]). *Suppose $\mathcal{F}$ is a family of graphs for which the function $\Psi_{\mathcal{F}}$ is not recursive. Then, the property of being $\mathcal{F}$-free cannot be uniformly tested with one-sided error.*

Note, that in Definition 6.1 the tester is defined as one that may have arbitrarily large query complexity, as long as it is a function of ϵ only. Hence, if $\Psi_{\mathcal{F}}$ is not recursive Theorem 6.5 rules out the possibility of designing a tester that receives ϵ as part of the input, no matter how large its query complexity is, as long as it is a function of ϵ only.

The main idea behind the proof of Theorem 6.5 is that by "inspecting" the behavior of a property tester for the property of being $\mathcal{F}$-free one can compute the function $\Psi_{\mathcal{F}}$. The main combinatorial tool in the proof of Theorem 6.5 is a Theorem of Erdős [Erd64] in extremal graph theory, which can be considered as a hypergraph version of Zarankiewicz problem [KST54]. As an immediate corollary of Theorems 4.1 and 6.5 we obtain the following result, which *precisely* characterizes the families of graphs $\mathcal{F}$, for which the property of being $\mathcal{F}$-free can be tested uniformly (recall that by Theorem 4.1, for *any* family $\mathcal{F}$, being $\mathcal{F}$-free is non-uniformly testable).

Corollary 6.6. *For every family of graphs $\mathcal{F}$, the property of being $\mathcal{F}$-free is uniformly testable with one-sided error if and only if the function $\Psi_{\mathcal{F}}$ is recursive.*

As for any monotone graph property $\mathcal{P}$, there is a family of graphs $\mathcal{F}$, for which $\mathcal{P}$ is equivalent to being $\mathcal{F}$-free (see Section 4) the above result also gives a *precisely* characterization of the monotone graph properties that can be tested uniformly.

An immediate consequence of Theorems 4.1 and 6.5 is that in order to separate uniform testing with one-sided error from non-uniform testing with one-sided error, and thus (almost) prove Theorem 6.3, it is enough to construct a family of graphs $\mathcal{F}$ with the following two properties: (i) There is an algorithm for deciding whether a graph F belongs to $\mathcal{F}$ (recall that we confine ourselves to decidable graph properties). (ii) The function $\Psi_{\mathcal{F}}$ is non-recursive. The main combinatorial ingredient in the construction of $\mathcal{F}$ is the fundamental theorem of Erdős [Erd59] in extremal graph theory, which guarantees the existence of graphs with arbitrarily large girth and chromatic number. As we want to prove Theorem 6.3 with a graph property, which is not only decidable, but even belongs to $coNP$, we need explicit constructions of such graphs. To this end, we use explicit constructions of expanders, which are given in [LPS88]. For the construction we also apply some ideas from the theory of recursive functions. Finally, in order to obtain that being $\mathcal{F}$-free cannot be tested even with two-sided error, we use a result of the first author ([GT03] Appendix D) about testing hereditary graph properties. The full details appear in [AS05c].

A final remark with regards to Theorem 6.3 is that as noted above, the testers considered in this theorem are not given the size of the input graph. It can be shown that if an oblivious and non-oblivious testers are given the size of the input then the notions of uniform and non-uniform testers coincide, namely, a graph property is uniformly testable if and only if it is non-uniformly testable. This observation combined with Theorem 6.3 can be considered as showing that a tester can use the size of the input in order to perform natural tasks, which it cannot do if it is not given this information.

7 Testing Hereditary Graph Properties

In this section we consider a family of graph properties, which significantly extends the family of monotone graph properties considered in Section 4. Besides including many additional interesting graph properties, this family can be proved to contain all the "natural" graph properties that can be tested with one-sided error. A sketch of the details follows.

A graph property is *hereditary* if it is closed under removal of vertices (and not necessarily under removal of edges). Clearly, every monotone graph property is also hereditary, but there are also many well-studied hereditary properties, which are not monotone. Examples are being Perfect, Chordal, Interval, Comparability, Permutation and more. See [Gol80], [MM99] and [RR01] for definitions and results about these as well as about other hereditary properties. The main results discussed in the previous sections deal with special cases of hereditary properties, namely, monotone properties. All the results about monotone properties heavily rely on the fact that removing an edge from a graph cannot increase its distance from satisfying a monotone property (see, e.g., Lemma 3.4). As it turns out, handling hereditary non-monotone graph properties, such as being Perfect or not containing an induced cycle of length 4, is significantly more involved than handling monotone properties. In Section 3 we have demonstrated that it is easy to show that for any *finite* $\mathcal{F}$ the property of being $\mathcal{F}$-free is testable. It is far more complicated to show that even for any single graph H, the property of being *induced* H-free is testable. This was only proved in [AFKS00] by applying Lemma 3.7.

For a (possibly infinite) family of graphs $\mathcal{F}$, a graph G is said to be *induced* $\mathcal{F}$-free if it contains no $F \in \mathcal{F}$ as an induced subgraph. As in the case of monotone properties, one can use essentially the same argument used in Lemma 3.5 in order to reduce the task of testing the property of being induced $\mathcal{F}$-free to a purely combinatorial problem.

Lemma 7.1. *Let $\mathcal{F}$ be a (possibly infinite) family of graphs, and suppose there are functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ such that the following holds for every $\epsilon > 0$: Every graph G on n vertices, which is ϵ-far from being* **induced** *$\mathcal{F}$-free contains at least $\delta_{\mathcal{F}}(\epsilon)n^f$* **induced** *copies of a graph $F \in \mathcal{F}$ of size $f \leq f_{\mathcal{F}}(\epsilon)$. Then, being induced $\mathcal{F}$-free is testable with one-sided error.*

The main result we discuss in this section is the following:

Theorem 7.2 ([AS05d]). *For any (possibly infinite) family of graph $\mathcal{F}$ there are functions $f_{\mathcal{F}}(\epsilon)$ and $\delta_{\mathcal{F}}(\epsilon)$ satisfying the conditions of Lemma 7.1.*

As we have done for monotone properties, one can define for any hereditary property $\mathcal{P}$, a (possibly infinite) family of graphs $\mathcal{F}_{\mathcal{P}}$ such that satisfying $\mathcal{P}$ is equivalent to being *induced* $\mathcal{F}_{\mathcal{P}}$-free. We simply put a graph F in $\mathcal{F}_{\mathcal{P}}$ if and only if F does not satisfy $\mathcal{P}$ but any graph obtained from F by removing a vertex satisfies $\mathcal{P}$. It thus follows that Theorem 7.2, combined with Lemma 7.1, implies the following

Theorem 7.3 ([AS05d]). *Every hereditary graph property is testable with one-sided error.*

It should be noted that besides certain partition properties such as having a large cut and having a large clique, which are testable with two-sided error by Theorem 2.1, essentially any graph property that was studied in the literature is hereditary. Thus, Theorem 7.3 combined with the graph partition problems of [GGR98] imply the testability of (nearly) any natural graph property.

The upper bounds for testing a given hereditary property, which Theorem 7.2 guarantees, have an enormous dependency on $1/\epsilon$. Moreover, the proof of the general result is quite involved. It may thus be interesting to address the following problem.

Open Problem 7.4. Find efficient and simple to analyze testers for specific hereditary properties such as being Perfect, Chordal and Interval.

One can infer from Theorem 7.2 the following analogue of Corollary 4.4:

Corollary 7.5 ([AS05d]). *For every hereditary graph property $\mathcal{P}$, there is a function $W_{\mathcal{P}}(\epsilon)$ with the following property: If G is ϵ-far from satisfying $\mathcal{P}$, then G contains an induced subgraph of size at most $W_{\mathcal{P}}(\epsilon)$, which does not satisfy $\mathcal{P}$.*

The above corollary implies, for example, that for every ϵ there is $c = c(\epsilon)$, such that if a graph G is ϵ-far from being Chordal then G contains an **induced** cycle of length at most c, and that similar results hold for any other hereditary property. This is non-trivial as it is not clear a priori that there is no graph that is, say, $\frac{1}{100}$-far from being Chordal and yet contains only induced cycles of length at least, say, $\Omega(\log n)$.

Another interesting application of Theorem 7.2 is an analogue of Theorem 4.6:

Theorem 7.6 ([AS05d]). *For any (possibly infinite) set of hereditary graph properties $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2, \ldots\}$, there is a function $\delta_{\mathcal{P}} : (0,1) \mapsto (0,1)$ with the following property: If a graph G is ϵ-far from satisfying all the properties of $\mathcal{P}$, then for some i, the graph G is $\delta_{\mathcal{P}}(\epsilon)$-far from satisfying $\mathcal{P}_i$.*

Observe, that when considering hereditary non-monotone properties, it is not even clear that if a graph is ϵ-far from satisfying *two* such properties $\mathcal{P} = \{\mathcal{P}_1, \mathcal{P}_2\}$, then it is $\delta_{\mathcal{P}}(\epsilon)$-far from satisfying one of them. (Recall that if the two properties are monotone then this is trivially true, as we can take $\delta_{\mathcal{P}}(\epsilon) = \epsilon/2$ for any such $\mathcal{P}$.) The above theorem asserts that this statement is true even for an infinite family of properties.

We now turn to discuss the most interesting application of Theorem 7.3. The problem of characterizing the graph properties, which are testable (either with one-sided or two-sided error) is considered by many to be the most important open problem in the area of property-testing. An interesting special case of this open problem is to characterize the graph properties, which are testable with one-sided error. This problem should be somewhat easier to resolve as numerous previous works, many of which are presented in this paper, demonstrated that testing with one-sided error is intimately related to various well-studied combinatorial problems, which can be handled using combinatorial tools. It seems, though, that even this seemingly easier problem is still very challenging. Using Theorem 7.3 one can obtain a characterization of the "natural" graph properties, which are testable with one-sided error.

As stated in Lemma 8.3 (see Section 8), by a result of [AFKS00] and [GT03], it is possible to assume that a property tester works by making its queries non-adaptively. In other words, the tester first picks a random subset of vertices S, and then continues without making additional queries. The following more restricted type of tester was introduced in [AS05d]:

Definition 7.7 (Oblivious Tester). *A tester (one-sided or two-sided) is said to be* oblivious *if it works as follows: given ϵ the tester computes an integer $Q = Q(\epsilon)$ and asks an oracle for a subgraph induced by a set of vertices S of size Q, where the oracle chooses S randomly and uniformly from the vertices of the input graph. If Q is larger than the size of the input graph then the oracle returns the entire graph. The tester then accepts or rejects according to the graph induced by S.*

In some sense, oblivious testers capture the essence of property testing as essentially all the testers that have been analyzed in the literature were in fact oblivious, or could trivially be turned into oblivious testers. Even the testers for properties such as having an independent set of size $\frac{1}{2}n$ or a cut with $\frac{1}{8}n^2$ edges (see [GGR98]), whose definition involves the size of the graph, have oblivious testers. The reason is simply that these properties can easily be expressed without using the size of the graph. Clearly, some properties cannot have oblivious testers, however, these properties are not natural. One example out of many, is the property of not containing an induced cycle of length 4 if the number of vertices is even, and not containing an induced cycle of length 5 if the number of vertices is odd. Using Theorem 7.3 it can be shown that if one considers only oblivious testers, then it is possible to precisely characterize the graph properties, which are testable with one-sided error. To state this characterization we need the following definition:

Definition 7.8 (Semi-Hereditary). *A graph property $\mathcal{P}$ is semi-hereditary if there exists a hereditary graph property $\mathcal{H}$ such that the following holds:*

1. *Any graph satisfying $\mathcal{P}$ also satisfies $\mathcal{H}$.*
2. *For any $\epsilon >$ there is an $M(\epsilon)$, such that any graph of size at least $M(\epsilon)$, which is ϵ-far from satisfying $\mathcal{P}$, contains an induced subgraph, which does not satisfy $\mathcal{H}$.*

Clearly, any hereditary graph property $\mathcal{P}$ is also semi-hereditary because we can take $\mathcal{H}$ in the above definition to be $\mathcal{P}$ itself. In simple words, a semi-hereditary $\mathcal{P}$ is obtained by taking a hereditary graph property $\mathcal{H}$, and removing from it a (possibly infinite, carefully chosen) set of graphs. This means that the first item in Definition 7.8 is satisfied. As there are graphs not satisfying $\mathcal{P}$ that do satisfy $\mathcal{H}$ these graphs do not contain any induced subgraph that does not satisfy $\mathcal{H}$ (because $\mathcal{H}$ is hereditary). The only restriction, which is needed in order to get item 2 in Definition 7.8, is that $\mathcal{P}$ will be such that for any $\epsilon > 0$ there will be only finitely many graphs that are ϵ-far from satisfying it, and yet contain no induced subgraph that does not satisfy $\mathcal{H}$. We are now ready to state the characterization.

Theorem 7.9 ([AS05d]). *A graph property $\mathcal{P}$ has an oblivious one-sided tester* **if and only if** *$\mathcal{P}$ is semi-hereditary.*

We briefly discuss the proof of Theorem 7.2, which is far more involved than the proof of Theorem 4.1. Again, the proof heavily relies on a graph functional similar to $\Psi_{\mathcal{F}}$. However, this time $\Psi_{\mathcal{F}}$ is defined with respect to the following new type of homomorphism:

Definition 7.10 (Colored-Homomorphism). *Let K be a complete graph whose vertices are colored black or white, and whose edges are colored black, white or grey (neither the vertex coloring nor the edge coloring is assumed to be proper in the standard sense). A* colored-homomorphism *from a graph F to a graph K is a mapping $\varphi : V(F) \mapsto V(K)$, which satisfies the following:*

1. *If $(u, v) \in E(F)$ then either $\varphi(u) = \varphi(v) = t$ and t is colored black, or $\varphi(u) \neq \varphi(v)$ and $(\varphi(u), \varphi(v))$ is colored black or grey.*
2. *If $(u, v) \notin E(F)$ then either $\varphi(u) = \varphi(v) = t$ and t is colored white, or $\varphi(u) \neq \varphi(v)$ and $(\varphi(u), \varphi(v))$ is colored white or grey.*

We briefly mention that the color black represents edges, the color white represents non-edges and grey edges represent "don't care" namely we allow edges and non-edges. Similarly, a black vertex represents a complete graph, while a white vertex represents an edgeless graph. The reason for using colored-homomorphism instead of standard homomorphism is that the fact that $H \mapsto K$ does not supply enough information about H. In particular, it does not supply any information about the non-edges of H. When dealing with monotone properties this information is not that important as we never increase the distance to satisfying a monotone property if we remove an edge from a graph. For non-monotone properties this is no longer the case, and this is where the notion of colored homomorphism comes in handy. The reader is referred to [AS05d] for the complete proof of Theorem 7.2. As noted by one of the referees, a notion similar to colored-homomorphism was investigated in [THKM06+] under the name M-partitions, see also [HN04] Section 5.7.

We conclude this section with some open problems. Two graph properties $\mathcal{P}_1$ and $\mathcal{P}_2$ are defined in [AFKS00] to be *indistinguishable* if for every $\epsilon > 0$ and large enough n, any graph on n vertices satisfying one property is never ϵ-far from satisfying the other. It is shown in [AFKS00] that in this case, $\mathcal{P}_1$ is testable if and only if $\mathcal{P}_2$ is testable. This suggests the following problem.

Open Problem 7.11. Characterize (either combinatorially, logically or by other means) the graph properties that are indistinguishable from some hereditary graph property.

Note, that by Theorem 7.3 and the result of [AFKS00] mentioned above, any property that is indistinguishable from a hereditary property is testable, possibly with two-sided error.

The general characterization of the testable graph properties seems a very challenging open problem. In fact, even characterizing graph properties that are testable with one-sided error seems a difficult task. As Theorem 7.9 demonstrates, if one considers only oblivious testers, then one can characterize the graph properties, which are testable with one-sided error. The following problem may thus be an interesting step towards a characterization of the testable graph properties.

Open Problem 7.12. Which graph properties have (possibly two-sided) oblivious testers, whose query complexity is bounded by a function of ϵ only?

Another intriguing problem is to prove a version of Theorem 5.8 for the family of hereditary graph properties. In fact, even the following special case seems hard to resolve.

Open Problem 7.13. For which graphs H can one compute in polynomial time the number of edge modifications needed to make a graph induced H-free?

8 Testing Subgraphs in Directed and Undirected Graphs

The main results discussed in the previous sections establish the testability of any hereditary graph property. However, checking the details of the proofs of these results reveal that the upper bounds, which these results supply, have an enormous dependency on $1/\epsilon$. Even for simple properties, these bounds are given by the so called WOW function, which is a tower of towers of exponents of height polynomial in $1/\epsilon$. This raises the natural problem of obtaining better upper bounds for specific properties. Specifically, the following is an intriguing open problem:

Open Problem 8.1. Which hereditary properties $\mathcal{P}$ can be tested with number of queries, which has a polynomial dependency on $1/\epsilon$.

In fact, even characterizing the monotone graph properties that are testable with $poly(1/\epsilon)$ queries ($poly(1/\epsilon) \equiv$ some polynomial in $1/\epsilon$) seems a challenging open problem. An easy example of a property that can be tested with $poly(1/\epsilon)$ queries is the property of being H-free when H is a single edge: It is easy to see that in this case a sample of $O(1/\epsilon)$ vertices (or pairs) suffices (and is also necessary). A more involved argument shows that the property of being k-colorable can also be tested with $poly(1/\epsilon)$ queries. See [GGR98] and [AK02]. A natural family of properties for which one may try to characterize the graph properties that are easily testable are the properties of being H-free and induced H-free, for a given fixed graph H. We will focus in this section on properties of being H-free, and only briefly discuss results about properties of being induced H-free at the end of this section. In this section we will denote by $\mathcal{P}_H$ the property of being H-free. We will say that $\mathcal{P}_H$ is *easily testable* with one sided error if it has a one-sided error tester with query complexity $poly(1/\epsilon)$. As we would like to focus on the combinatorial part of the problem, we will restate the problem combinatorially in Corollary 8.5 below, for which we need the following two lemmas:

Lemma 8.2 ([AS05c]). *Let T be a one-sided error tester for testing $\mathcal{P}_H$. Then, T must accept an input graph if the graph induced by the sample of vertices satisfies $\mathcal{P}_H$.*

Lemma 8.3 ([AFKS00],[GT03]). *If there exists an ϵ-tester for a graph property that makes q queries, then there exists such an ϵ-tester that makes its queries by uniformly and randomly choosing a set of $2q$ vertices and querying all their pairs. In particular, it is a non-adaptive ϵ-tester making $\binom{2q}{2}$ queries.*

Lemma 8.4. *Let H be a fixed graph on h vertices. Then, $\mathcal{P}_H$ is testable with one sided error and with query complexity $(q(\epsilon))^{O(1)}$ if and only if any graph on n vertices, which is ϵ-far from being H-free, contains at least $n^h/q^c(\epsilon)$ copies of H, for some $c = c(h)$.*

Proof. Suppose first that any graph on n vertices, which is ϵ-far from being H-free contains at least $n^h/q^c(\epsilon)$ copies of H. The tester for $\mathcal{P}_H$ samples a set of vertices S, of size $h \cdot q^c(\epsilon)$ and accepts the input graph if and only if S spans no copy of H. The tester is clearly one-sided. Also, if the input graph is ϵ-far from being H-free then by assumption, each h-tuple of vertices spans a copy of H with probability at least $1/q^c(\epsilon)$. Thus, a sample of size $3h \cdot q^c(\epsilon)$ spans a copy of H with probability at least $2/3$. Suppose now that for any (small enough) ϵ and any (large enough) n there is a graph G of size n, which is ϵ-far from being H-free, and yet it contains only $n^h/q^h(\epsilon)$ copies of H. Thus, by the union bound, the probability that a random set of $q(\epsilon)$ vertices spans a copy of H is at most $h!\binom{q(\epsilon)}{h}/q^h(\epsilon) < 2/3$. By Lemmas 8.3 and 8.2 this means that $\mathcal{P}_H$ cannot be tested with one-sided error and query-complexity $q(\epsilon)$. ☐

As our focus in this section is in testing with number of queries, which is polynomial in $1/\epsilon$, the following implication of the above will be useful.

Corollary 8.5. *$\mathcal{P}_H$ is easily testable with one sided error if and only if any graph on n vertices, which is ϵ-far from being H-free, contains at least $n^h/poly(1/\epsilon)$ copies of H.*

The problem of characterizing the graphs H for which $\mathcal{P}_H$ is easily testable with one-sided error was resolved in [Alon02].

Theorem 8.6 ([Alon02]). *$\mathcal{P}_H$ is easily testable with one-sided error if and only if H is bipartite.*

The first part of the above characterization, that of showing that when H is bipartite, then $poly(1/\epsilon)$ queries suffice for testing $\mathcal{P}_H$, follows by showing that any graph with ϵn^2 edges, contains at least $\epsilon^{h^2/4} n^h$ copies of H. As any graph, which is ϵ-far from being H-free, must contain at least ϵn^2 edges, one gets from Lemma 8.5 that in this case being H-free is testable with $poly(1/\epsilon)$ queries. The proof of the other direction is more complicated. To prove this, one shows that for any non-bipartite graph H there are graphs, which are ϵ-far from being H-free and yet contain only $\epsilon^{c \log 1/\epsilon} n^h$ copies of H. By Lemma 8.5 this is sufficient for obtaining the lower bound. The construction of [Alon02] involves a subtle application of a variant of Behrend's construction [Beh46] of a dense subset of the first n integers containing no 3-term arithmetic progressions. It also applies some properties of cores of graphs defined in subsection 1.2. Theorem 8.6 was extended in [AS04a] to the case of two-sided error testers:

Theorem 8.7 ([AS04a]). *$\mathcal{P}_H$ is easily testable (either with one-sided or two-sided error) if and only if H is bipartite.*

One of the main results of [AS04a], is a characterizing of the *directed* graphs (digraphs, for short) H, for which $\mathcal{P}_H$ is easily testable. As it turns out, the case of directed graphs is significantly more involved than the case of undirected graphs. To state the main result of [AS04a] recall the definition of a core of a graph (or a digraph) given in subsection 1.2. The following is the main result of [AS04a]:

Theorem 8.8 ([AS04a]). *Let H be a fixed connected digraph, and let K be its core. Then $\mathcal{P}_H$ is easily testable if and only if K is either a cycle of length 2 or any oriented tree.*

Observe, that Theorem 8.7 follows as a special case of Theorem 8.8, but with the following (equivalent) formulation: $\mathcal{P}_H$ is easily testable if and only if the core of H is a single edge. As is apparent from the statement of Theorem 8.8, the characterization for digraphs is far more complicated than the characterization for undirected graphs, which is given in Theorem 8.7. The characterization for undirected graphs is also simple in the sense that one can check it in polynomial time. As it turns out, the characterization for digraphs is not complicated by chance as the following holds:

Proposition 8.9. *The following problem is NP-complete: Given a digraph H, decide if $\mathcal{P}_H$ is easily testable.*

The above proposition follows easily by combining Theorem 8.8 with a result of of Hell, Nešetřil, and Zhu [HNZ96], who proved that deciding if the core of a digraph is a tree is NP-complete.

Observe, that Theorem 8.8 implies that the property $\mathcal{P}_C$ is easily testable for the oriented cycle C on the vertices $v_1, \ldots, v_{2k}$, that consists of two edge-disjoint directed paths from v_1 to v_{k+1}, as each of the two paths is a core of C. Theorem 8.8 also implies that the property $\mathcal{P}_{C'}$ is *not* easily testable for every oriented cycle C' that is obtained from the above cycle C, by changing the direction of *any* single edge, because in this case the core of C' is the entire digraph. This example shows that the query complexity needed to test $\mathcal{P}_H$ does not depend solely on the structure of H as an undirected graph.

The proof of Theorem 8.8 has three major steps. The simplest involves showing that if the core of H is a 2-cycle then $\mathcal{P}_H$ is easily testable. In this case it is shown that a variant of the argument used in [Alon02] in order to show that testing $\mathcal{P}_H$ is easily testable for any bipartite H, gives an analogous result for directed H, whose core is a 2-cycle. In fact, the argument in [AS04a], when considered for the special case of undirected graphs, improves the upper bound given in [Alon02]. The second step involves the case when the core of H is neither a tree nor a 2-cycle. In this case one has to construct graphs, which are ϵ-far from being H-free and yet contain few copies of H. The construction of such graphs also uses Behrend's construction. The most involved case to handle is in showing that when the core of H is a tree then $\mathcal{P}_H$ is also easily testable.

In this section we have only considered properties of being H-free, because of their relation to homomorphisms. The problem of testing properties of being *induced* H-free have also been considered, but in this case the characterization is much simpler (to state, not to prove) as in this case for any graph or digraph on at least 5 vertices, it can be shown that induced H-freeness cannot be tested with *poly*$(1/\epsilon)$ queries. See the proof in [AS04b].

A natural extension of the results considered in this section is the case of k-uniform hypergraphs (k-graphs, for short). In this case a k-graph G is ϵ-far from being (induced) H-free if one must add/remove at least ϵn^k edges in order to turn G into a (induced) H-free k-graph. The case of being induced H-free was essentially resolved in [AS05a] where it was shown that for $k \geq 3$, the only H for which the property of being induced H-free can be tested with *poly*$(1/\epsilon)$ queries are the single k-edges (which are trivially testable with $O(1/\epsilon)$ queries) as well as a single 3-graph on 4 vertices, H, for which it is still not known if being induced H-free can be tested with *poly*$(1/\epsilon)$ queries. The properties of being (not necessarily induced) H-free, denoted as above by $\mathcal{P}_H$, seems much more involved. It is easy to show (see [KNR02] and [AS05a]) that for any k-partite k-graph H, the property $\mathcal{P}_H$ is testable with *poly*$(1/\epsilon)$ queries. [AS05a] also contains a proof of a sufficient condition for inferring that for a given H, $\mathcal{P}_H$ is not testable with *poly*$(1/\epsilon)$ queries. Regretfully, this condition does not include all non-k-partite k-graphs. It will be interesting to complete the picture by proving (or disproving) the following:

Open Problem 8.10. Is it the case that for $k \geq 3$, $\mathcal{P}_H$ is testable with *poly*$(1/\epsilon)$ queries if and only if H is a k-partite k-graph.

References

[Alon02] N. Alon, Testing subgraphs in large graphs, *Proc.* 42^{nd} *IEEE FOCS*, IEEE (2001), 434–441. Also: *Random Structures and Algorithms* 21 (2002), 359–370.

[Alon06$^+$] N. Alon, Ranking tournaments, to appear.

[ADLRY94] N. Alon, R. A. Duke, H. Lefmann, V. Rödl and R. Yuster, The algorithmic aspects of the Regularity Lemma, *Proc.* 33^{rd} *IEEE FOCS, Pittsburgh*, IEEE (1992), 473–481. Also: *J. of Algorithms* 16 (1994), 80–109.

[AFKS00] N. Alon, E. Fischer, M. Krivelevich and M. Szegedy, Efficient testing of large graphs, *Proc. of* 40^{th} *FOCS, New York, NY*, IEEE (1999), 656–666. Also, *Combinatorica* 20 (2000), 451–476.

[AK02] N. Alon and M. Krivelevich, Testing k-colorability, *SIAM J. Discrete Math.* 15 (2002), 211–227.

[AS03] N. Alon and A. Shapira, Testing satisfiability, *Proc.* 13^{th} *Annual ACM-SIAM SODA*, ACM Press (2002), 645–654. Also: *Journal of Algorithms* 47 (2003), 87–103.

[AS04a] N. Alon and A. Shapira, Testing subgraphs in directed graphs, *Proc. of the* 35^{th} *Annual Symp. on Theory of Computing (STOC), San Diego, California, 2003*, 700–709. Also: *JCSS* 69 (2004), 354–382.

[AS04b] N. Alon and A. Shapira, A characterization of easily testable induced subgraphs, *Proc. of the* 15^{th} *Annual ACM-SIAM SODA*, ACM Press (2004), 935–944. Also: *Combinatorics, Probability and Computing*, to appear.

[AS05a] N. Alon and A. Shapira, Linear equation, arithmetic progressions and hypergraph property testing, *Proc. of the* 16^{th} *Annual ACM-SIAM SODA, ACM Press (2005)*, 708–717.

[AS05b] N. Alon and A. Shapira, Every monotone graph property is testable, *Proc. of the* 37^{th} *Annual Symp. on Theory of Computing (STOC), Baltimore, Maryland*, 2005, 128–137.

[AS05c] N. Alon and A. Shapira, A separation theorem in property-testing, manuscript, 2005.

[AS05d] N. Alon and A. Shapira, A characterization of the (natural) graph properties testable with one-sided error, *Proc. of FOCS 2005*, 429–438.

[ASS05] N. Alon, A. Shapira and B. Sudakov, Additive approximation for edge-deletion problems, *Proc. of FOCS 2005*, 419–428.

[AS00] N. Alon and J. H. Spencer, *The Probabilistic Method*, Second Edition, Wiley, New York, 2000.

[BBW06$^+$] J. Balogh, B. Bollobás and D. Weinreich, Measures on monotone properties of graphs, *Discrete Applied Mathematics*, to appear.

[BKS06$^+$] J. Balogh, P. Keevash and B. Sudakov, On the minimal degree implying equality of the largest triangle-free and bipartite subgraphs, submitted.

[BH90] J. Bang-Jensen and P. Hell, The effect of two cycles on the complexity of colorings by directed graphs, *Discrete Applied Math.* 26 (1990), 1–23.

[Beh46] F. A. Behrend, On sets of integers which contain no three terms in arithmetic progression, *Proc. National Academy of Sciences USA* 32 (1946), 331–332.

[BLR93] M. Blum, M. Luby and R. Rubinfeld, Self-testing/correcting with applications to numerical problems, *JCSS* 47 (1993), 549–595.

[BSTT06+] J. Bondy, J. Shen, S. Thomassé and C. Thomassen, Density conditions for triangles in multipartite graphs, *Combinatorica*, to appear.

[CMNR97] K. Cirino, S. Muthukrishnan, N. Narayanaswamy and H. Ramesh, graph editing to bipartite interval graphs: exact and asymptotic bounds, *Proc. of* 17^{th} *FSTTCS* (1997), 37–53.

[EM00] D. Eichhorn and D. Mubayi, Edge-coloring cliques with many colors on subcliques, *Combinatorica* 20 (2000), 441–444.

[EC88] E. S. El-Mallah and C. J. Colbourn, The complexity of some edge-deletion problems, *IEEE transactions on circuits and systems*, 35 (1988), 354–362.

[Erd59] P. Erdős, Graph theory and probability, *Canad. J. Mathematics* 11 (1959), 34–38.

[Erd64] P. Erdős, On extremal problems of graphs and generalized graphs, *Israel J. Math.* 2 (1964), 183–190.

[EFR86] P. Erdős, P. Frankl and V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs Combin.* 2 (1986) 113–121.

[EG97] P. Erdős and A. Gyárfás, A variant of the classical Ramsey problem, *Combinatorica* 17 (1997), 459–467.

[THKM06+] T. Feder, P. Hell, S. Klein, and R. Motwani, Complexity of list partitions, *Proc. of STOC 1999*, 464–472. Also, *SIAM J. Comput.*, in press.

[Fish01] E. Fischer, The art of uninformed decisions: A primer to property testing, *The Computational Complexity Column of The Bulletin of the European Association for Theoretical Computer Science* 75 (2001), 97–126.

[FN05] E. Fischer and I. Newman, Testing versus estimation of graph properties, *Proc. of the* 37^{th} *Annual Symp. on Theory of Computing (STOC), Baltimore, Maryland, 2005*, 138–146.

[FK96] E. Friedgut and G. Kalai, Every monotone graph property has a sharp threshold, *Proc. Amer. Math. Soc.* 124 (1996), 2993–3002.

[GJ79] M.R. Garey and D.S. Johnson, *Computers and Intratability: A guide to the Theory of NP-Completeness*, W. H. Freeman and Co., San Francisco, 1979.

[GGKS95] P.W. Goldberg, M.C. Golumbic, H. Kaplan and R. Shamir, Four strikes against physical mapping of DNA, *Journal of Computational Biology* 2 (1995), 139–152.

[Gol98] O. Goldreich, Combinatorial property testing – a survey, In: *Randomization Methods in Algorithm Design* (P. Pardalos, S. Rajasekaran and J. Rolim eds.), AMS-DIMACS (1998), 45–60.

[GGR98] O. Goldreich, S. Goldwasser and D. Ron, Property testing and its connection to learning and approximation, *Proc. of* 37^{th} *Annual IEEE FOCS*, (1996), 339–348. Also: *JACM* 45(4): 653–750 (1998).

[GR97] O. Goldreich and D. Ron, Property Testing in Bounded-Degree Graphs, *Proc. of STOC 1997*, 406–415.

[GT03] O. Goldreich and L. Trevisan, Three theorems regarding testing graph properties, *Proc.* 42^{nd} *IEEE FOCS*, IEEE (2001), 460–469. Also, *Random Structures and Algorithms* 23(1):23-57, 2003.

[Gol80] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, 1980.

[GKS94] M. C. Golumbic, H. Kaplan and R. Shamir, On the complexity of DNA physical mapping, *Advances in Applied Mathematics*, 15 (1994), 251–261.

[GRS90] R. L. Graham, B. L. Rothschild and J. H. Spencer, *Ramsey Theory*, Second Edition, Wiley, New York, 1990.

[Gow06$^+$] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, manuscript.

[HN04] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford University Press, 2004.

[HN92] P. Hell and J. Nešetřil, The core of a graph, *Discrete Math* 109 (1992), 117–126.

[HNZ96] P. Hell, J. Nešetřil, and X. Zhu, Duality of graph homomorphisms, in: *Combinatorics, Paul Erdős is Eighty*, (D. Miklós et. al, eds.), Bolyai Society Mathematical Studies, Vol.2, 1996, pp. 271–282.

[KMS98] D. Karger, R. Motwani and M. Sudan, Approximate graph coloring by semidefinite programming, *JACM* 45(2), 1998, 246–265.

[KNR02] Y. Kohayakawa, B. Nagle and V. Rödl, Efficient testing of hypergraphs,*Proc. of* 29^{th} *ICALP*, (2002), 1017–1028.

[KS96] J. Komlós and M. Simonovits, Szemerédi's Regularity Lemma and its applications in graph theory. In: *Combinatorics, Paul Erdős is Eighty*, Vol II (D. Miklós, V. T. Sós, T. Szönyi eds.), János Bolyai Math. Soc., Budapest (1996), 295–352.

[KST54] T. Kővári, V. T. Sós and P. Turán, On a problem of K. Zarankiewicz, *Colloquium Math.* 3 (1954), 50–57.

[LY80] J. Lewis and M. Yannakakis, The node deletion problem for hereditary properties is NP-complete, *JCSS* 20 (1980), 219–230.

[Lov79] L. Lovász, On the shannon capacity of a graph, *IEEE Transactions on Information Theory* 25(1), 1979, 1–7.

[LPS88] A. Lubotzky, R. Phillips and P. Sarnak, Ramanujan graphs, *Combinatorica* 8 (1988), 261–277.

[MM99] T. A. McKee and F. R. McMorris, *Topics in Intersection Graph Theory*, SIAM, Philadelphia, PA, 1999.

[NRS06$^+$] B. Nagle, V. Rödl and M. Schacht, The counting lemma for regular k-uniform hypegraphs, manuscript.

[PR02] M. Parnas and D. Ron, Testing the diameter of graphs, *Random structures and algorithms* 20 (2002), 165–183.

[Pap94] C. Papadimitriou, *Computational Complexity*, Addison Wesley, 1994.

[PRR04] M. Parnas, D. Ron and R. Rubinfeld, Tolerant property testing and distance approximation, manuscrip, 2004.

[RR01] J. L. Ramírez-Alfonsín, B. A. Reed (Editors), *Perfect Graphs*, Wiley, 2001.

[RD85] V. Rödl and R. Duke, On graphs with small subgraphs of large chromatic number, *Graphs and Combinatorics* 1 (1985), 91–96.

[RS04] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, *Random Structures and Algorithms* 25 (2004), 1–42.

[Ron01] D. Ron, Property testing, in: P. M. Pardalos, S. Rajasekaran, J. Reif and J. D. P. Rolim, editors, *Handbook of Randomized Computing*, Vol. II, Kluwer Academic Publishers, 2001, 597–649.

[Rose72] J. D. Rose, A graph-theoretic study of the numerical solution of sparse positive-definite systems of linear equations, *Graph Theory and Computing*, R. C. Reed, ed., Academic Press, N.Y., 1972, 183–217.

[RS96] R. Rubinfeld and M. Sudan, Robust characterization of polynomials with applications to program testing, *SIAM J. on Computing* 25 (1996), 252–271.

[Sze78] E. Szemerédi, Regular partitions of graphs, In: *Proc. Colloque Inter. CNRS* (J. C. Bermond, J. C. Fournier, M. Las Vergnas and D. Sotteau, eds.), 1978, 399–401.

[West01] D. B. West, *Introduction to Graph Theory*, Prentice Hall, 2001.

[Xue94] J. Xue, Edge-maximal triangulated subgraphs and heuristics for the maximum clique problem. Networks 24 (1994), 109–120.

Counting Graph Homomorphisms

Christian Borgs[1], Jennifer Chayes[2], László Lovász[3], Vera T. Sós[4*], and Katalin Vesztergombi[5]

[1] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
borgs@microsoft.com
[2] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
jchayes@microsoft.com
[3] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
lovasz@microsoft.com
[4] Alfréd Rényi Institute of Mathematics, Budapest, Reáltanoda U. 13-15, H-1053 Budapest, Hungary
sos@renyi.hu
[5] Department of Computer Science, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary
veszterk@cs.elte.hu

Summary. Counting homomorphisms between graphs (often with weights) comes up in a wide variety of areas, including extremal graph theory, properties of graph products, partition functions in statistical physics and property testing of large graphs.

In this paper we survey recent developments in the study of homomorphism numbers, including the characterization of the homomorphism numbers in terms of the semidefiniteness of "connection matrices", and some applications of this fact in extremal graph theory.

We define a distance of two graphs in terms of similarity of their global structure, which also reflects the closeness of (appropriately scaled) homomorphism numbers into the two graphs. We use homomorphism numbers to define convergence of a sequence of graphs, and show that a graph sequence is convergent if and only if it is Cauchy in this distance. Every convergent graph sequence has a limit in the form of a symmetric measurable function in two variables. We use these notions of distance and graph limits to give a general theory for parameter testing.

The convergence can also be characterized in terms of mappings of the graphs into fixed small graphs, which is strongly connected to important parameters like ground state energy in statistical physics, and to weighted maximum cut problems in computer science.

AMS Subject Classification. Primary 05C99, Secondary 05C35, 05C80, 82B99.
Keywords. Graph homomorphism, partition function, extremal graphs, Szemerédi Regularity Lemma, connection matrix, convergent graph sequence, limits of graphs.

* Research supported in part by OTKA grants No. TO32236, TO39210, TO42750.

1 Introduction

For two finite graphs G and H, let $\hom(G,H)$ denote the number of homomorphisms (adjacency-preserving mappings) from G to H. Counting homomorphisms between graphs has many interesting aspects.

(a) A large part of extremal graph theory can be expressed as inequalities between various homomorphism numbers. For example, Turán's Theorem for triangles follows from the inequality (due to Goodman [Goo59]):

$$\hom(K_1,H)\hom(K_3,H) \ge \hom(K_2,H)(2\hom(K_2,H) - \hom(K_1,H)^2). \quad (1)$$

One may wish to obtain a characterization of such inequalities.

(b) Homomorphism numbers characterize graphs: it was proved in [Lov67] that if two graphs G and G' have the property that $\hom(F,G) = \hom(F,G')$ for every finite graph F, then G and G' are isomorphic (one actually needs one additional condition, the condition that both graphs are twin-free; see Section 2.1 for the definition of this notion). In other words, let us order all finite graphs in a sequence $(F_1, F_2 \ldots)$, and assign to each graph G its *profile*, the (infinite) sequence $(\hom(F_1,G), \hom(F_2,G)\ldots)$; then this sequence characterizes G. (This fact can be used to prove, for example, the cancellation property of strong multiplication of graphs).

It is often worthwhile to normalize the homomorphism numbers, and consider the *homomorphism densities*

$$t(F,G) = \frac{\hom(F,G)}{|V(G)|^{|V(F)|}}. \quad (2)$$

(Thus $t(F,G)$ is the probability that a random map of $V(F)$ into $V(G)$ is a homomorphism.) Instead of the profile $(\hom(F_1,G), \hom(F_2,G), \ldots)$, we can consider the *scaled profile* $(t(F_1,G), t(F_2,G), \ldots)$ of the graph G. Such a normalization was first introduced in [ELS79].

(c) Partition functions of many models in statistical mechanics can be expressed as graph homomorphism functions. For example, let G be an $n \times n$ grid, and suppose that every node of G (every "site") can be in one of two states, "UP" or "DOWN". The properties of the system are such that no two adjacent sites can be "UP". A "configuration" is a valid assignment of states to each node. The number of configurations is the number of independent sets of nodes in G, which in turn can be expressed as the number of homomorphisms of G into the graph H consisting of two nodes, "UP" and "DOWN", connected by an edge, and with an additional loop at "DOWN". To define the thermodynamic functions in physical models, one needs to extend the notion of graph homomorphism to the case when the nodes and edges of H have weights (see Section 2.1).

(d) Suppose that G is a huge graph, and we know the numbers $\hom(F,G)$ (exactly or approximately) for small graphs F. What kind of information can be derived about the global structure of G? The long-standing Reconstruction Conjecture is equivalent to the assertion that it is enough to know all numbers $\hom(F,G)$ with $|V(F)| < |V(G)|$ in order to recover the isomorphism type of G. The fact that this is unsolved shows the difficulty of this kind of question, but our interest here is in the case when much less is given: $\hom(F,G)$ is only known for very small graphs F.

(e) This is closely related to an important area of computer science called *Property Testing*. In this model, we have a huge graph about which we can obtain information only by taking a small sample of the nodes and examine the subgraph induced by them. This is equivalent to knowing the homomorphism densities (2) for graphs F of small size. What makes the theory of Property Testing interesting is the fact that from such meager local information nontrivial properties and parameters of the graph can be inferred.

(f) Increasing sequences of (sparse) graphs generated by some specific random rule of growth have recently been used to model the Internet; see [BR03] and references therein. What are the limiting properties of such graph sequences? To what extent can these properties be derived from local observation (observing a neighborhood of bounded radius of a few randomly chosen nodes? This question can be rephrased as follows: To what extent are these properties characterized by the homomorphism numbers of smaller graphs into the modeling sequence?

The main setup of our studies is the following. If we are given a (large, usually simple) graph G, we may try to study its local structure by counting the homomorphisms of various "small" graphs F into G; and we can study its global structure by counting its homomorphisms into various small graphs H (called "softcore" weighted graphs; see Section 2.1 for a definition of softcore). So the scheme to keep in mind is

$$F \longrightarrow G \longrightarrow H. \tag{3}$$

According to the above discussion, in the scheme (3), the study of the local structure of G by "probing from the left with F" is related to property testing, while the study of the global structure of G by "probing from the right with H" is related to statistical physics. As in statistical physics, the best choice of graphs H to "probe G from the right" is not simple unweighted graphs but weighted graphs. Furthermore, besides counting (weighted) homomorphisms into H, it is also useful to consider maximizing the weight of such homomorphisms, which is again related to well-studied questions both in statistical physics and graph property testing.

The number $\hom(F,G)$, as a function of F with G fixed, is a graph parameter (a function of graphs F invariant under isomorphism). Graph parameters arising this way were characterized in [FLS04]. The necessary and sufficient condition involves certain matrices, called *connection matrices*, associated with the graph parameter (see Section 3 for the definition): These matrices must be positive definite and must satisfy a rank condition. The semidefiniteness condition is in fact familiar from statistical physics, where it is called *reflection positivity.*

The scaled profile does not determine the graph: if we "blow up" every node into the same number of "twins", then we get a graph with exactly the same scaled profile. It turns out that this is all: any two graphs with the same scaled profile are obtained from one and the same graph by blowing up its nodes in two different ways.

Now we come to our main question: What can be said about two graphs whose scaled profiles are approximately the same? Which properties of a graph G are determined if we only know a few of the numbers $\hom(F, G)$, and even these are only known approximately? This question turns out to be very interesting and it leads to a number of results, and even more open problems, leading to quasirandom and generalized quasirandom graphs, and connecting to the "Property Testing" research in computer science and to statistical physics.

There is a way to measure the "distance" of two graphs so that they are close in this distance if and only if they have approximately the same scaled profile [BCLSV06$^+$]. This distance has many nice properties. On the one hand, important parameters like the triangle density or the fraction of edges in the maximum cut are continuous (often even Lipschitz) functions in this metric. On the other hand, a sufficiently large random subgraph of an arbitrarily large graph will be close to the whole graph with large probability. This fact explains several results in the theory of Property Testing. Szemerédi's Regularity Lemma (at least in its weaker but more effective form due to Frieze and Kannan [FK99]) can be rephrased as follows: for every $\varepsilon > 0$, all graphs with at most $2^{2/\varepsilon^2}$ nodes form an ε-net in the metric space of all graphs. In our context, an ε-net is defined as a set of weighted graphs such that, for every graph G, there exists a graph H in the set which has at most distance ε from G.

Once we make the set of all graphs into a metric space, we can make it complete. Is there any combinatorial meaning of the new points in the completion? Surprisingly, the answer is in the affirmative, and in fact in more than one way [LSz04, LSz05a]. These limit points can be described as symmetric measurable functions $W : [0,1]^2 \to [0,1]$ modulo measure-preserving transformations, as reflection positive graph parameters, as random graph models satisfying some natural compatibility conditions, or as probability distributions of countable graphs with natural symmetries (see Section 4).

It is an important property of this completion that it is compact, so that every infinite sequence of graphs has a convergent subsequence. Several arguments in extremal graph theory and elsewhere can be simplified by going to the limit and thereby getting rid of remainder terms.

We conclude this introduction by mentioning two related bodies of work. There are many interesting question that concern the existence of graph homomorphisms rather than their number (an example is 4-colorability of a graph), and such questions have been studied quite extensively, especially by the Czech school. These results are described in the recent book by Hell and Nešetřil [HN04]. The set of all homomorphisms between two graphs can be endowed with a topological structure, which turns out to be an important tool in the study of chromatic number. See the book of Matoušek [Mat03], and also the recent papers of Babson and Kozlov [BK06+a, BK06+b].

This paper is organized as follows. In Section 2 we define homomorphism numbers, including homomorphisms into a measurable function, and describe the basic examples. In Section 3 we define connection matrices, study their rank, and use their semidefiniteness to characterize homomorphism functions. We also describe an analogous (but much more difficult) edge-coloring version. In Section 4 we define convergence of a sequence of dense graphs, and show that they have interesting limit objects, which can be described as measurable functions, reflection positive graph parameters, or very natural models of finite or countable random graphs. In Section 5, we introduce a metric on graphs that corresponds to the above notion of convergence, and show its connection with Szemerédi's Regularity Lemma and graph property testing. Section 6 studies homomorphisms from large graphs, rather than into large graphs, which leads to quantities of both combinatorial and physical interest; we show that they too can be used to characterize convergent graph sequences. Section 7 contains some applications to extremal graph theory. In Section 8 we conclude with describing some analogous (but less complete) results for sequences of graphs with bounded degree.

2 Homomorphism Numbers

2.1 Unweighted and Weighted Graphs

A graph is *simple* if it has no loops or parallel edges. A *graph parameter* is a function defined on finite graphs, invariant under isomorphisms. We'll talk of a *simple graph parameter* if it is only defined on simple graphs. Sometimes it is convenient to think of a simple graph parameter as a function defined on all graphs with multiple edges (but no loops) that is invariant under adding parallel edges. A graph parameter t is called multiplicative if $t(G) = t(G_1)t(G_2)$ whenever G is the disjoint union of G_1 and G_2. We say that a graph parameter is *normalized* if its value on K_1, the graph with one node and no edge, is 1. (Note that if a graph parameter is multiplicative and not identically 0, then its value on K_0, the graph with no nodes and edges, is 1. The graph parameter $t(\cdot, G)$ introduced in the introduction is multiplicative and normalized for every graph G.

Recall that for two (finite) simple graphs F and G, $\hom(F, G)$ denotes the number of homomorphisms (adjacency preserving maps) from F to G.

A *weighted graph* G is a graph with a weight $\alpha_i(G)$ associated with each node i and a weight $\beta_{ij}(G)$ associated with each edge ij. We'll assume (unless otherwise stated) that the weights $\alpha_i(G)$ are positive. The weights $\beta_{ij}(G)$ will be real, and most often nonnegative. If the graph G is understood from the context, we will also use the notation α_i and β_{ij}.

An edge with weight 0 will play the same role as no edge between those nodes, so we could assume that we only consider weighted complete graphs with loops at all nodes (but this is not always convenient). A weighted graph is called *softcore* if it is a complete graph with loops at each node, and every edgeweight is strictly positive. An *unweighted graph* is a weighted graph where all the nodeweights and edgeweights are 1.

For a weighted graph G, we denote by $\alpha(G)$ the sum of its nodeweights. Often it will be useful to divide all nodeweights by $\alpha(G)$, to get a weighted graph $\widehat{G}$ in which the sum of nodeweights is 1.

Let F and G be two weighted graphs. To every map $\phi: V(F) \to V(G)$, we assign the weight

$$\hom_\phi(F,G) = \prod_{uv\in E(F)} \left[\beta_{\phi(u)\phi(v)}(G)\right]^{\beta_{uv}(F)} \tag{4}$$

(here $0^0 = 1$). We then define

$$\hom(F,G) = \sum_{\phi:\ V(F)\to V(G)} \alpha_\phi \hom_\phi(F,G), \tag{5}$$

where

$$\alpha_\phi = \prod_{u\in V(F)} \left[\alpha_{\phi(u)}(G)\right]^{\alpha_u(F)}. \tag{6}$$

(A little care is necessary, since the exponential $\left[\beta_{\phi(u)\phi(v)}(G)\right]^{\beta_{uv}(F)}$ may not be well defined; but it well defined e.g. if the edge weights are positive. This problem will not arise in the cases we consider.)

We'll use this definition most often in the case when F is a simple unweighted graph, so that

$$\alpha_\phi = \prod_{u\in V(F)} \alpha_{\phi(u)}(G)$$

and

$$\hom_\phi(F,G) = \prod_{uv\in E(F)} \beta_{\phi(u)\phi(v)}(G).$$

An interesting case is when, in addition, $\alpha(G) = 1$. Then the nodeweights in G define a probability distribution on $V(G)$, and the coefficient α_ϕ is the probability of the random map ϕ (if the images of the nodes of F are chosen independently). So $\hom(F,G)$ is the expectation of $\hom_\phi(F,G)$.

We can extend homomorphism densities (2) to the case when G is a weighted graph with nodeweights α_i and edgeweights β_{ij} by replacing $|V(G)|$ by $\alpha(G)$:

$$t(F,G) = \frac{\hom(F,G)}{\alpha(G)^{|V(F)|}} = \hom(F,\widehat{G}).$$

Let $\mathcal{F} = \{F_1, F_2, \ldots\}$ denote the set of (isomorphism types of) all simple finite graphs. To every weighted graph G, we assign its *hom-profile* (or briefly *profile*), the (infinite) vector

$$h_G = (\hom(F_1,G), \hom(F_2,G), \ldots) \in \mathbb{R}^{\mathcal{F}}.$$

Recall that we define the *scaled profile* of G as the (infinite) vector $t_G = (t(F_1,G), t(F_2,G), \ldots) \in \mathbb{R}^{\mathcal{F}}$. The scaled profile does not determine the graph. For example, if G is an unweighted graph and G' is obtained from G by replacing every node by N independent nodes, then

$$t_{G'} = t_G.$$

More generally, let G be a weighted graph and let $u, v \in V(G)$ be *twins*, i.e., $\beta_{uw}(G) = \beta_{vw}(G)$ for every $w \in V(G)$ (note that $\alpha_u(G)$ may be different from $\alpha_v(G)$). Merging twins in a weighted graph does not change its scaled profile. Furthermore, if we multiply all nodeweights of a weighted graph by the same positive constant, then its scaled profile does not change. If we merge twins as long as we can, we say that we have performed *twin-reduction.*

Proposition 2.1 ([Lov06$^+$]). *If two weighted graphs have the same scaled profile, then after twin-reduction one can be obtained from the other by multiplying all nodeweights by the same positive scalar.*

One of our main concerns will be: *if we only know a bounded number of entries of the scaled profile of a graph G, and even this only approximately, to what degree is the graph determined?*

From a graph-theoretic perspective, the following variations of homomorphism functions are perhaps more important: Let $\mathrm{inj}(F,G)$ denote the number of homomorphisms that are injective on the nodes, and $\mathrm{ind}(F,G)$, the number of embeddings as an induced subgraph. Finally, let $\mathrm{surj}(F,G)$ denote the number of homomorphisms that are surjective on the nodes.

For weighted graphs, $\mathrm{inj}(F,G)$ and $\mathrm{surj}(F,G)$ are easily defined by restricting the sum in (5) to sums over injective and surjective maps, respectively, but the definition of $\mathrm{ind}(F,G)$ requires some care. Here we only consider the case where F is simple, and G is a weighted graph without loops. We then define

$$\mathrm{ind}(F,G) = \sum_{\phi:\ V(F)\to V(G)} \alpha_\phi \mathrm{ind}_\phi(F,G), \tag{7}$$

where the sum goes over all injective maps from $V(F)$ to $V(G)$ and

$$\mathrm{ind}_\phi(F,G) = \prod_{uv\in E(F)} \beta_{\phi(u)\phi(v)}(G) \prod_{uv\in E(\overline{F})} (1 - \beta_{\phi(u)\phi(v)}(G)), \tag{8}$$

with $\overline{F}$ denoting the complement of F (again defined to be a simple graph).

Analogously to the hom-profile, we can define the *inj-profile* and *ind-profile* of a graph G. It follows from the identities to be discussed below that any of the profiles determines the others. It is obvious that the ind-profile determines the graph (look at the largest non-zero entry). It follows that the inj-profile and hom-profile of a graph also determine the graph (up to isomorphism) [Lov67]. For future reference, we also define the set of simple graph parameters which is the pointwise closure of the set of all hom-profiles:

$$\mathcal{T}_0 = \{t(\cdot) : \exists\, (G_n) \text{ s.t. } t(F) = \lim_{n\to\infty} t(F, G_n) \quad \forall \,\text{finite}\, F\}.$$

Here we restrict (G_n) to be a sequence of simple graphs. The reason for the subscript on $\mathcal{T}$ is that, in our paper [BCLSV06$^+$], we will consider a more general class of graph parameters, which will be denoted by $\mathcal{T}$. Note also that in [LSz04], the authors used $\mathcal{T}$ rather than $\mathcal{T}_0$ to denote the smaller class considered here.

2.2 Simple Properties

There are some simple identities that hold for homomorphism numbers. If F is the disjoint union of two graphs F_1 and F_2, then

$$\hom(F, G) = \hom(F_1, G)\hom(F_2, G). \tag{9}$$

If F is connected, and G is the disjoint union of two graphs G_1 and G_2,then

$$\hom(F, G) = \hom(F, G_1) + \hom(F, G_2). \tag{10}$$

So in a sense it is enough to study homomorphisms between connected graphs.

For two simple graphs G_1, G_2, we define their *categorial product* $G_1 \times G_2$ as the graph with node set $V(G_1) \times V(G_2)$, in which (i_1, i_2) is connected to (j_1, j_2) $(i_1, j_1 \in V(G_1),\ i_2, j_2 \in V(G_2))$ if and only if $i_1 j_1 \in E(G_1)$ and $i_2 j_2 \in E(G_2)$. The definition can be extended to weighted graphs (possibly with loops) by defining the weight of the node (i_1, i_2) as the product of the weights of the nodes i_1 and i_2, and the weight of the edge $(i_1, i_2)(j_1, j_2)$ as the product of the weights of the edges $i_1 j_1$ and $i_2 j_2$. For this product, we have the identity

$$\hom(F, G_1 \times G_2) = \hom(F, G_1) \cdot \hom(F, G_2). \tag{11}$$

What about $\hom(F_1 \times F_2, G)$? There is no identity for this number in terms of the notions introduced so far, but there is one if one also introduces the operation of exponentiation (which we do not discuss here; see [Lov67]).

For a fixed graph G, identity (9) gives an algebraic relation between the entries of its hom-profile. It was proved by Whitney [Whi32] that there are no other algebraic relations between these entries valid for all graphs G. A slightly stronger result was proved in [ELS79]: *the projection of $\mathcal{T}_0$ to the coordinates corresponding to any finite set of connected graphs is full-dimensional.* This excludes any other kind of equations (e.g. exponential) between these numbers.

There are simple identities relating homomorphism numbers with the injective and induced versions. In order to spare the reader from separate provisos for each relation, we restrict ourselves to the case where both F and G are simple graphs. If Θ is any equivalence relation on $V(F)$, then we denote by F/Θ the graph obtained by identifying nodes that belong to the same class of Θ. Note that this may create loops and parallel edges.

We have some easy relations:

$$\hom(F,G) = \sum_{\Theta} \mathrm{inj}(F/\Theta, G) \tag{12}$$

and

$$\mathrm{inj}(F,G) = \sum_{F' \supset F} \mathrm{ind}(F', G), \tag{13}$$

where the sum runs over graphs $F' \supset F$ with the same node set.

From these, we can get reverse relations by Möbius inversion (or inclusion-exclusion):

$$\mathrm{ind}(F,G) = \sum_{F' \supset F} (-1)^{|E(F') \setminus E(F)|} \mathrm{inj}(F', G), \tag{14}$$

and

$$\mathrm{inj}(F,G) = \sum_{\Theta} \mu(\Theta) \hom(F/\Theta, G), \tag{15}$$

where the last sum runs over equivalence relations and

$$\mu(\Theta) = \prod_{A \in \Theta}^{k} \Big((-1)^{(|A|-1)} (|A|-1)! \Big),$$

with the product running over all classes $A \in \Theta$.

We have the following relations describing complementation (as an operation from simple graphs to simple graphs):

$$\mathrm{ind}(F, \overline{G}) = \mathrm{ind}(\overline{F}, G), \tag{16}$$

$$\mathrm{inj}(F, \overline{G}) = \sum_{F' \subset F} (-1)^{|E(F')|} \mathrm{inj}(F', G) \tag{17}$$

and

$$\hom(F, \overline{G}) = \sum_{F' \subset F} (-1)^{|E(F')|} \hom(F', G). \tag{18}$$

2.3 Examples of Homomorphism Functions

Example 2.2 (Stars and degrees). Let S_k denote the star with k nodes. Then for any graph G on n nodes,

$$\hom(S_k, G) = \sum_{i=1}^{n} d_i^{k-1}, \tag{19}$$

where $d_1, \dots, d_n$ are the degrees of G. Hence $\hom(S_k, G)^{1/(k-1)}$ tends to the maximum degree of G as $k \to \infty$.

Example 2.3 (Cycles and eigenvalues). Let C_k denote the cycle on k nodes, and again let G be any graph on n nodes. Then

$$\hom(C_k, G) = \sum_{i=1}^{n} \lambda_i^k, \tag{20}$$

where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of the adjacency matrix of G. Hence $\hom(C_{2k}, G)^{1/(2k)}$ tends to the largest eigenvalue of G as $k \to \infty$.

Example 2.4 (Independent sets). Let H be the graph on two nodes, with an edge connecting the two nodes and a loop at one of the nodes. Then for every simple graph G, $\hom(G, H)$ is the number of independent sets of nodes in G.

Example 2.5 (Colorings). It is easy to see that $\hom(G, K_q)$ is the number of colorings of the graph G with q colors. It is well known that for a fixed G, this number is a polynomial in q, called the *chromatic polynomial.* The chromatic polynomial defines a graph invariant for every complex number q, but this cannot be expressed as the number of homomorphisms into any graph unless q is a nonnegative integer [FLW, FLS04] (cf. Example 3.4).

It is often useful to consider homomorphisms into a fixed graph H as generalized colorings, where the colors are the nodes of H, and every edge of H imposes a constraint on the coloring that these two colors cannot be used at adjacent nodes.

Example 2.6 (Maximum cut). Let H denote the looped complete graph on two nodes, weighted as follows: the non-loop edge has weight 2; all other edges and nodes have weight 1. Then for every simple graph G with n nodes,

$$\log_2 \hom(G, H) - n \le \mathrm{MaxCut}(G) \le \log_2 \hom(G, H).$$

where $\mathrm{MaxCut}(G)$ denotes the size of the maximum cut in G. So unless G is very sparse, $\log_2 \hom(G, H)$ is a good approximation of the maximum cut in G.

Example 2.7 (Random graphs). Let $G = G(n,p)$ be a random graph with n nodes and edge-density p. Then for every simple graph F with k nodes,

$$\mathsf{E}(\hom(F,G)) = (1+o(1))n^k p^{|E(F)|} \qquad (n \to \infty).$$

By a straightforward application of high concentration results, it follows that $\hom(F,G)$ is very close to its expectation with large probability.

Example 2.8 (Partition functions of the Ising model). Let G be any simple graph, and let $T > 0$, $h \geq 0$, and J be three real parameters. Let H be the looped complete graph on two nodes, denoted by $+$ and $-$, weighted as follows: $\alpha_+ = e^{h/T}$, $\alpha_- = e^{-h/T}$, $\beta_{++} = \beta_{--}$, $\beta_{+-} = \beta_{-+}$, and $\beta_{++}/\beta_{+-} = e^{2J/T}$. Then $\hom(G,H)$ is the partition function of the Ising model on the graph G at temperature T with coupling J in external magnetic field h.

2.4 Homomorphisms into Measurable Functions

The following definition from [FLS04] (which will play an important role later on) generalizes the homomorphism function. Every bounded function $W : [0,1]^2 \to \mathbb{R}$ defines a graph parameter as follows: For a finite graph F on k nodes, let

$$t(F,W) = \int_{[0,1]^k} \prod_{ij\in E(F)} W(x_i,x_j)\, dx_1 \, \ldots \, dx_k.$$

(We can think of the interval $[0,1]$ as the set of nodes, and of the value $W(x,y)$ as the weight of the edge xy.) While this definition is meaningful for all graphs F, we will mostly use it for simple graphs.

It is easy to see that for every weighted graph G, the graph parameter $t(\cdot,G)$ is a special case. We may assume that $V(G) = \{1,\ldots,n\}$ and $\alpha(G) = 1$. Define a function $W_G : [0,1]^2 \to [0,1]$ as follows. For $(x,y) \in [0,1]^2$, let a and b be determined by

$$\alpha_1(G) + \ldots + \alpha_{a-1}(G) \leq x < \alpha_1(G) + \ldots + \alpha_a(G),$$
$$\alpha_1(G) + \ldots + \alpha_{b-1}(G) \leq y < \alpha_1(G) + \ldots + \alpha_b(G),$$

and let

$$W_G(x,y) = \beta_{ab}(G).$$

(Informally, W_G is obtained by replacing the (i,j) entry in the weighted adjacency matrix of G by a rectangle of size $\alpha_i \times \alpha_j$, and define the function value on this square as β_{ij}.) Then

$$t(F,G) = t(F,W_G)$$

for every finite simple graph G.

Example 2.9. For an undirected simple graph F, let $\mathsf{eul}(F)$ denote the number of eulerian orientations of F (i.e., orientations in which every node has the same outdegree as indegree). By Euler's theorem, $\mathsf{eul}(F) = 0$ if and only if F has a node with odd degree.

It can be shown [LSz04] that this graph parameter can be represented in the form $t(\cdot, W)$, where

$$W(x, y) = 2\cos(2\pi(x - y)).$$

On the other hand, it follows e.g. from Theorem 3.6 below that eul is not of the form $\hom(\cdot, G)$ with any finite weighted graph G.

3 Connection Matrices

3.1 The Connection Matrix of a Graph Parameter

A *k-labeled graph* ($k \geq 0$) is a finite graph in which k nodes are labeled by $1, 2, \dots k$. Two k-labeled graphs are *isomorphic*, if there is a label-preserving isomorphism between them. We denote by K_k the k-labeled complete graph on k-nodes, and by O_k, the k-labeled graph on k nodes with no edges.

Let G_1 and G_2 be two k-labeled graphs. Their *product* G_1G_2 is defined as follows: we take their disjoint union, and then identify nodes with the same label. Clearly this multiplication is associative and commutative. For 0-labeled graphs, this notation is in line with our notation for disjoint union.

The following construction is central to the theory of homomorphisms functions. Let f be any graph parameter. For every integer $k \geq 0$, we define the following (infinite) matrix $M(f, k)$. The rows and columns are indexed by isomorphism types of k-labeled graphs. The entry in the intersection of the row corresponding to G_1 and the column corresponding to G_2 is $f(G_1G_2)$. We call the matrices $M(f, k)$ the *connection matrices* of the graph parameter f (see Figure 1).

For a simple graph parameter, the above construction causes trouble if we get multiple edges when gluing the two graphs. In this case, we suppress the edge multiplicities in G_1G_2 when defining the entry corresponding to the pair (G_1, G_2).

3.2 The Rank of Connection Matrices

Connection matrices of a graph parameter are infinite matrices and their rank may be infinite. However, the rank is quite often finite, and if so, this fact has interesting consequences. Let us denote by $\mathrm{rk}(f, k)$ the rank of the k-th connection matrix of the graph parameter f.

We start with several examples of graph parameters for which the rank of connection matrices is finite. The most important case for us will be when the graph parameter is defined as $\hom(\cdot, H)$ for some fixed weighted graph H. This will be discussed in detail in the next section.

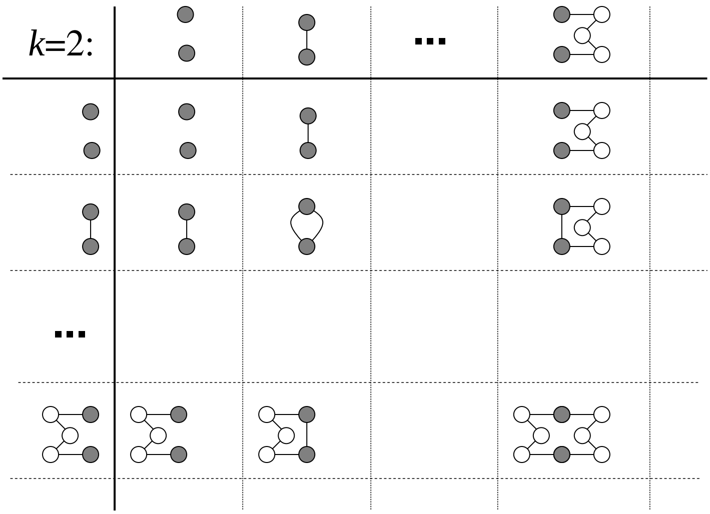


Fig. 1. A small part of the connection matrix for $k = 2$. The matrix entries are obtained by evaluating the parameter on the graph shown

Example 3.1 (Edges). Let $\mathsf{e}(G) = |E(G)|$ denote the number of edges in G. Then $\mathsf{e}(G_1G_2) = \mathsf{e}(G_1) + \mathsf{e}(G_2)$, and so $M(\mathsf{e}, k)$ is the sum of two matrices of rank 1. Thus $M(\mathsf{e}, k)$ has rank 2, so $\mathrm{rk}(\mathsf{e}, k) = 2$ for all k.

If we restrict $\mathsf{e}(G)$ to simple graphs G to get a simple graph parameter e', then the situation is more complicated: we have

$$\mathsf{e}'(G_1G_2) = \mathsf{e}'(G_1) + \mathsf{e}'(G_2) - \mathsf{e}'(G_1 \cap G_2).$$

Rewriting $\mathsf{e}'(G_1 \cap G_2)$ as $x(G_1)^{\mathsf{T}}x(G_2)$ where $x(G)$ is the $\binom{k}{2}$-dimensional vector with entries $x_{ij}(G) = 1$ if G contains an edge joining the labeled vertices i and j and $x_{ij}(G) = 0$ otherwise, we see that the matrix whose (G_1, G_2) entry is $\mathsf{e}'(G_1 \cap G_2)$ has rank $\binom{k}{2}$, implying that $\mathrm{rk}(\mathsf{e}', k) \leq \binom{k}{2} + 2$. One can check that this is the exact value.

Example 3.2 (Subgraphs). Let $\mathsf{subg}(G)$ denote the number of spanning subgraphs of G, i.e., $\mathsf{subg}(G) = 2^{\mathsf{e}(G)}$. Then $\mathsf{subg}(G_1G_2) = \mathsf{subg}(G_1)\mathsf{subg}(G_2)$, and so $M(\mathsf{subg}, k)$ has rank 1. Thus $\mathrm{rk}(\mathsf{subg}, k) = 1$ for all k.

Again, the version when we only consider simple graphs is more complicated: Let $\mathsf{subg}'(G)$ denote this simple graph parameter. Then

$$\mathsf{subg}'(G_1G_2) = \frac{\mathsf{subg}'(G_1)\mathsf{subg}'(G_2)}{\mathsf{subg}'(G_1 \cap G_2)}.$$

The first two factors do not change the rank, and the rows of the matrix given by the second factor are determined by the edges induced by the labeled nodes, so it has only $2^{\binom{k}{2}}$ different rows, implying that $\mathrm{rk}(\mathsf{subg}', k) \leq 2^{\binom{k}{2}}$. Again one can check that this is the exact value.

Example 3.3 (Matchings). Let $\mathsf{pmatch}(G)$ denote the number of perfect matchings in the graph G. It is trivial that $\mathsf{pmatch}(G)$ is multiplicative. We claim that its node-rank-connectivity is exponentially bounded:

$$r_{\mathsf{pmatch}}(k) \le 2^k.$$

Let G be a k-labeled graph, let $X \subseteq [k] = \{1, \dots, k\}$, and let $\mathsf{pmatch}(G, X)$ denote the number of matchings in G that match all the unlabeled nodes and the nodes with label in X, but not any of the other labeled nodes. Then we have for any two k-labeled graphs G_1, G_2

$$\mathsf{pmatch}(G_1 G_2) = \sum_{X_1 \cap X_2 = \emptyset,\, X_1 \cup X_2 = [k]} \mathsf{pmatch}(G_1, X_1)\mathsf{pmatch}(G_2, X_2).$$

This can be read as follows: The matrix $M(\mathsf{pmatch}, k)$ can be written as a product $N^{\mathsf{T}} W N$, where N has infinitely many rows indexed by k-labeled graphs, but only 2^k columns, indexed by subsets of $[k]$,

$$N_{G,X} = \mathsf{pmatch}(G, X),$$

and W is a symmetric $2^k \times 2^k$ matrix, where

$$W_{X_1, X_2} = \begin{cases} 1 \text{ if } X_1 = [k] \setminus X_2, \\ 0 \text{ otherwise.} \end{cases}$$

Hence the rank of $M(\mathsf{pmatch}, k)$ is at most 2^k (it is not hard to see that in fact equality holds).

Example 3.4 (Chromatic polynomial). We have seen that the number of q-colorings is a special case of homomorphism functions. This number is the evaluation of the chromatic polynomial $\mathsf{chr}(G; x)$ at nonnegative integers q. What about evaluations at other values? It turns out that these evaluations violate both conditions in Theorem 3.6 below [FLW]. For every fixed x, this is a multiplicative graph parameter. To describe its rank-connectivity, we need the following notation. For $k, q \in \mathbb{Z}_+$, let $B_{k,q}$ denote the number of partitions of a k-element set into at most q parts. So $B_k = B_{k,k}$ is the k-th Bell number. With this notation,

$$\mathrm{rk}(\mathsf{chr}, k) = \begin{cases} B_{k,x} & \text{if } x \text{ is a positive integer,} \\ B_k & \text{otherwise.} \end{cases}$$

Note that this is always finite: if x is a positive integer, then it is bounded by x^k, but otherwise it grows faster than c^k for every c. Similar results can be derived for the Tutte polynomial, where the exceptional values are the hyperbolas in the Tutte plane for which $(x-1)(y-1)$ is a positive integer.

Let f be a graph parameter that is not identically 0. Then f is multiplicative if and only if $f(K_0) = 1$ (K_0 is the empty graph) and $\mathrm{rk}(f, 0) = 1$. Every multiplicative graph parameter f satisfies the inequality

$$\mathrm{rk}(f, k+l) \geq \mathrm{rk}(f, k) \cdot \mathrm{rk}(f, l). \tag{21}$$

In the most important special case (to be discussed below) when f is a homomorphism function, a stronger version of property (21) holds: the sequence $\mathrm{rk}(f, k)$ is logconvex. We do not know if this property holds for more general graph parameters.

Finiteness of the rank connectivity function has interesting algorithmic consequences:

Theorem 3.5 ([FLW]). *If $r(f, k)$ is finite for some k, then f can be computed in polynomial time for graphs with treewidth at most k.*

3.3 Connection Matrices of Homomorphisms

Homomorphism functions, which are our main concern, provide the most important class of graph parameters for which connection matrices have finite rank. In fact, connection matrices can be used to characterize these parameters, as the following theorem of Freedman, Lovász and Schrijver shows.

Theorem 3.6 ([FLS04]). *The graph parameter f, defined on graphs with multiple edges but no loops, is equal to* $\hom(\cdot, H)$ *for some weighted graph H on q nodes if and only if*

(a) *$M(f, k)$ is positive semidefinite and*
(b) $\mathrm{rk}(f, k) \leq q^k$ *for all k.*

In terms of statistical physics, this theorem can be viewed as a characterization of partition functions of models whose degrees of freedom sit on vertices (as opposed to the edge coloring models considered below). The property that $M(f, k)$ is positive semidefinite is related to the "reflection positivity" property in statistical physics, and we will call a graph parameter *reflection positive* if $M(f, k)$ is positive semidefinite for every k.

The proof of the necessity of the conditions in Theorem 3.6 is easy and it is instructive to present it here. (The sufficiency is more involved, and the proof is based on algebraic considerations.)

We need the following notation: For any k-labeled graph G and mapping $\phi: [k] \to V(H)$, let

$$\hom_\phi(G, H) = \sum_{\substack{\psi:\ V(G) \to V(H) \\ \psi \text{ extends } \phi}} \frac{\alpha_\psi}{\alpha_\phi} \hom_\psi(G, H), \tag{22}$$

so that

$$\hom(G, H) = \sum_{\phi:\ [k] \to V(H)} \alpha_\phi \hom_\phi(G, H). \tag{23}$$

For any two k-labeled graph G_1 and G_2,

$$\hom_\phi(G_1G_2, H) = \hom_\phi(G_1, H)\hom_\phi(G_2, H). \tag{24}$$

The decomposition (23) writes the matrix $M(f,k)$ as the sum of $|V(H)|^k$ matrices, one for each mapping $\phi : [k] \to V(H)$; (24) shows that these matrices are positive semidefinite and have rank 1.

In the presence of condition (a), condition (b) in Theorem 3.6 can be replaced by the following quite different type of condition. To formulate it, we need the notion of a quantum graph, defined as a formal linear combination of graphs with real coefficients; but the condition concerns only the existence of a single 2-labeled quantum graph [LSz05a]:

(c) *There is a 2-labeled quantum graph g_0 with the following property: if G is a 2-labeled graph having no edge between the labeled nodes, and G' denotes the graph obtained from G by identifying the two labeled nodes, then $f(g_0G) = f(G')$.*

In other words, attaching g_0 at two nodes is effectively the same as identifying the two nodes (this is why it is called a “contractor” in [LSz05a]).

Let us conclude with a discussion of the independence of the two conditions in Theorem 3.6.

In Example 3.3 we saw that the number of perfect matchings in a graph provides an example for a graph parameter for which the rank of connection matrices grows simply exponentially. This parameter is also multiplicative, so for $k = 0$ the connection matrix is positive semidefinite. But it is easy to see that for $k = 1$, the submatrix indexed by K_1 and K_2 is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which is not positive semidefinite. Thus the number of perfect matchings cannot be represented as a homomorphism function.

Recalling Examples 3.1 and 3.2, let us define the multigraph parameter f by $f(G) = 1/\mathsf{subg}'(G) = 2^{-e(G')}$, where G' is obtained from G by removing duplicate edges. As in Example 3.2, the rank of the connection matrix $M(f,k)$ grows as $2^{\binom{k}{2}}$. It is further not hard to check that $M(f,k)$ is positive semidefinite. The graph parameter f is, in fact, the limit of parameters of the form $\hom(\cdot, H)$: take homomorphisms into a random graph $H = \mathbf{G}(n, 1/2)$, with all nodeweights $1/n$ and all edge-weights 1. But the rank of its connection matrices is finite but superexponential, so the parameter is not of the form $\hom(\cdot, H)$.

This example also illustrates the importance of the condition that f is defined on graphs with multiple edges for the validity of 3.6. Indeed, for a simple graph G (*i.e.*, if G has no multiple edges), $f(G) = 2^{-e'(G)}$ can be represented as the number of homomorphisms into the graph $K_1(1/2)$, consisting of a single node with a loop, where the node has weight 1 and the loop has weight 1/2.

The chromatic polynomial (Example 3.4) was another example whose connection matrices had superexponential rank growth if the variable x was not a nonnegative integer. Here the reflection positivity condition gives the same condition on x: the k-th connection matrix is positive semidefinite if and only if either x is a positive integer or $k \leq x+1$. Thus $M(\mathsf{chr}, k)$ is semidefinite for all k if and only if x is a positive integer.

3.4 The Exact Rank of Connection Matrices for Homomorphisms

How good is the upper bound on the rank given in Theorem 3.6? It can be proved [Lov06$^+$] that equality holds in the "generic" case.

One reason why the upper bound is not always reached are twins. As remarked earlier, twin-reduction in H does not change the numbers $\hom(G, H)$, and so it does not change the connection matrices (but of course it decreases the upper bounds $|V(H)|^k$). So we may assume that H is twin-free.

The second reason for rank loss in connection matrices is that if H has a proper automorphism (a permutation of the nodes that preserves both the nodeweights and edgeweights), then in formula (22), any two terms defined by a mapping $\phi : [k] \to V(H)$ and $\phi\sigma$ ($\sigma \in \mathrm{Aut}(H)$) are equal, so the sum of all such terms is still rank 1. So the rank of $M(\hom(\cdot, H), k)$ is at most the number of orbits of the automorphism group of H on ordered k-tuples of its nodes.

Theorem 3.7 ([Lov06$^+$]). *Assume that the target graph H is twin-free. Then for every k, $\mathrm{rk}(\hom(\cdot, H), k)$ is the number of orbits of the automorphism group of H on ordered k-tuples of its nodes.*

It is worthwhile to formulate two corollaries.

Corollary 3.8. *Let H be a weighted graph that has no twins and no automorphisms. Then $\mathrm{rk}(\hom(\cdot, H), k) = |V(H)|^k$ for every k.*

Swapping twins i and j is "almost" an automorphism: the only additional condition needed is that $\alpha_i = \alpha_j$. In particular, for unweighted graphs the condition that there are no automorphisms implies that there are no twins.

Our second corollary is in fact an equivalent reformulation of Theorem 3.7 in the framework of quantum graphs. To state it, we need the following notation: given a weighted graph H, a k-labeled quantum graph x and nodes $i_1, \dots, i_k \in V(H)$, we define $\hom_{i_1 \dots i_k}(x, H)$ to be the number of homomorphisms from x to H such that the labeled nodes are mapped into $i_1 \dots i_k$.

Corollary 3.9. *Let H be a weighted graph that has no twins, and let $h : V(H)^k \to \mathbb{R}$. Then there exists a k-labeled quantum graph x such that $\hom_{i_1 \dots i_k}(x, H) = h(i_1, \dots, i_k)$ for all $i_1, \dots, i_k \in V(H)$ if and only if h is invariant under the automorphisms of H.*

3.5 Extensions: Directed Graphs, Hypergraphs, Semigroups

In Theorem 3.6 we allowed parallel edges in the graphs G, but no loops. Indeed, the representation theorem is false if G can have loops: it is not hard to check that the graph parameter

$$\mathsf{loop}(G) = 2^{-\#\mathrm{loops}}$$

cannot be represented as a homomorphism function, even though its connection matrix $M(\mathsf{loop}, k)$ is positive semidefinite and has rank 1. To get a representation theorem for graphs with loops, each loop e in the target graph H must have two weights: one which is used when a non-loop edge of G is mapped onto e, and the other, when a loop of G is mapped onto e. With this modification, the theorem remains valid.

The constructions and results above are in fact more general; they extend to directed graphs and hypergraphs. One new element in the case of directed graphs is the following. For a directed graph, homomorphism functions can be defined by the same formulas (4) and (5), except that weights are now assigned to the directed edges of G and H. But there are (at least) two substantially different ways to define connection matrices.

(1) The easier way is to define k-labeled digraphs similarly, and glue them together just like we did in the undirected case. The related theorem and proof are precisely the same as above.

(2) The alternative generalization of Theorem 3.6 to directed graphs is a bit more interesting. (Example 3.13 below is a case when this second theorem applies.) For this, we consider weighted directed graphs in which the edgeweights can be complex. Such a graph H is called *Hermitian* if for every arc uv with (complex) weight β_{uv}, the arc vu is also present and has weight $\beta_{vu} = \overline{\beta}_{uv}$.

For any directed graph D, let D^* be the digraph obtained from D by reversing all arcs. For two k-labeled digraphs D and D', let DD' denote, as before, their union with the labeled nodes identified. For any complex-valued digraph parameter f defined on loopless directed graphs and for each natural number k, we define the matrix $\tilde{M}(f,k)$ as follows: its rows and columns are indexed by k-labeled directed graphs, and the entry in position D_1, D_2 is $f(D_1^* D_2)$.

Theorem 3.10 ([LSc06$^+$]). *Let f be a complex valued digraph parameter. Then $f = \hom(\cdot, H)$ for some Hermitian weighted digraph H if and only if $f(K_0) = 1$ and there exists a $d \geq 0$ such that, for each $k \geq 0$, $\tilde{M}_k$ is positive semidefinite and has rank at most d^k.*

Note that $\tilde{M}_k$ is a complex valued matrix. The condition that it is positive semidefinite includes the condition that it is Hermitian.

There is a common formulation of these results, using semigroups; see [LSc06$^+$] for details.

3.6 Edge Coloring Models

Let G be a finite graph. An *edge coloring model* or *edge model* is determined by a finite set C and a mapping $h: \mathbb{Z}_+^C \to \mathbb{R}_+$, which we call the *node evaluation function.* Here C is the set of possible edge colors; for any coloring of the edges, we think of $h(a)$ as the value of a node incident with $a(c)$ edges with the color c ($c \in C$). In terms of statistical physics, an edge coloring is a state of the system, and $\log h(a)$ is the contribution of a node (incident with $a(c)$ edges with the color c) to the energy of the state.[1]

To be more precise, for an edge-coloring $\phi: E(G) \to C$ and node v, let $a_{\phi,v}(c)$ denote the number of edges e incident with v with $\phi(e) = c$. So $a_{\phi,v} \in \mathbb{Z}_+^C$ is the "local view" of node v. The weight of the assignment ϕ is defined by

$$w(\phi) = \prod_{v\in V(G)} h(a_{\phi,v}),$$

and the *edge coloring parameter*, by

$$\operatorname{col}(G,h) = \sum_{\phi:\ E(G)\to C} w(\phi).$$

(It will be also useful to allow a single edge with no endpoints; we call this graph the *circle*, and denote it by $\bigcirc$. By definition, $\operatorname{col}(\bigcirc, h) = |C|$.)

We can define edge-connection matrices that are analogous to the connection matrices defined before: Instead of gluing graphs together along nodes, we glue them together along edges. To be precise, we define a *k-broken graph* as a k-labeled graph in which the labeled nodes have degree one. (It is best to think of the labeled nodes not as nodes of the graph, but rather as points where the k edges sticking out of the rest of the graph are broken off.) We allow that both ends of an edge be broken off.

For two k-broken graphs G_1 and G_2, we define $G_1^*G_2$ by gluing together the corresponding broken ends of G_1 and G_2. These ends are not nodes of the resulting graph any more, so $G_1^*G_2$ is different from the graph G_1G_2 we would obtain by gluing together G_1 and G_2 as k-labeled graphs. One very important difference is that while G_1G_2 is k-labeled, $G_1^*G_2$ has no broken edges any more, and so it is not k-broken. This fact leads to considerable difficulties in the treatment of edge models.

For every graph parameter f and integer $k \geq 0$, we define the *edge-connection matrix* $M'(f,k)$ as follows. The rows and columns are indexed by isomorphism types of k-broken graphs. The entry in the intersection of the row corresponding to G_1 and the column corresponding to G_2 is $f(G_1^*G_2)$. Note that for $k = 0$, we have $M(f,0) = M'(f,0)$, but for other values of k, connection and edge-connection matrices are different. We say that f is *edge reflection positive*, if $M'(f,k)$ is positive semidefinite for every $k \geq 0$.

[1] For this reason, these models are usually called vertex models in the physics literature.

It is easy to see (similarly as in the case of homomorphism functions) that if $h: \mathbb{Z}_+^C \to \mathbb{R}_+$ and $f = \mathrm{col}(\cdot, h)$ is an edge-coloring function, then

$$\mathrm{rk}(M'(f,k)) \le |C|^k,$$

and $M'(f,k)$ is positive semidefinite. Unlike in the case of node-connection matrices, these two properties are not independent any more:

Proposition 3.11 ([Sze05]). *If f is a multiplicative graph parameter such that $M'(f,k)$ is positive semidefinite for every $k \ge 0$, then $f(\bigcirc)$ is a nonnegative integer and* $\mathrm{rk}(M'(f,k)) \le f(\bigcirc)^k$.

The analogue of Theorem 3.6 is even simpler to state (but much more difficult to prove:

Theorem 3.12 ([Sze05]). *A graph parameter f can be represented as $f(\cdot) = \mathrm{col}(\cdot, h)$ for some edge coloring model h if and only if it is multiplicative and edge-reflection positive.*

Just as for homomorphism functions, it is natural to ask what determines the rank of connection matrices of edge models. This question seems to lead to difficult algebraic questions in group representations, and is unanswered at this time.

3.7 Edge Colorings and Homomorphisms

The connection between homomorphism functions and edge coloring functions seems to go farther than analogy, but it is not well understood.

In one direction, edge coloring functions are more general than homomorphism functions. This connection is easy in the directed case. We can generalize the edge coloring model to directed graphs in the obvious way. It is easy to see that the directed edge model is more general than the directed homomorphism model: If we are given a pair $H = (a, B)$ ($a \in \mathbb{R}_+^q$, $B \in \mathbb{C}^{q\times q}$, with entries α_c and β_{uv}, resp.), then to every homomorphism $\phi: V(G) \to [q]$ we can assign an edge-coloring in which the edge ij is colored with the pair $\psi(ij) = (\phi(i), \phi(j))$. The evaluation function at a node v is given as follows: if there is a color c such that all the outgoing edges have the same first color c and all the incoming edges have the same second color c, then the value is

$$\alpha_c \prod_{u:\ uv\in E(G)} \beta_{\psi(uv)_1,\psi(uv)_2};$$

otherwise, the node evaluates to 0. It is easy to see that an edge-coloring has nonzero weight only if it comes from a homomorphism, and in that case, the weight of the edge-coloring is the same as the weight of the corresponding homomorphism.

It is not obvious, but it is true, that undirected edge coloring functions generalize undirected homomorphism functions [Sze05], at least if complex values are allowed for h. It is not clear which (real) homomorphism functions can be obtained as edge coloring functions with a real valued h.

In the opposite direction, edge coloring models cannot be translated into node coloring models (homomorphisms) in general; but there are some nontrivial examples of important graph parameters that are defined as edge coloring functions, but that can also be represented as homomorphism functions in a nontrivial way. A general understanding of these examples would be very interesting.

Example 3.13 (Nowhere-zero flows). Let $\mathsf{eul}(G) = 1$ if G is eulerian (i.e., all nodes have even degree), and $\mathsf{eul}(G) = 0$ otherwise. To represent this function as a homomorphism function, let

$$a = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

It was noted by de la Harpe and Jones [HJ93] that for the weighted graph $H = (a, B)$ we have $\hom(G, H) = \mathsf{eul}(G)$.

This example can be generalized quite a bit. Let Γ be a finite abelian group and let $S \subseteq \Gamma$ be such that S is closed under inversion. For any graph G, fix an orientation of the edges. An *S-flow* is an assignment of an element of S to each edge such that for each node v, the product of elements assigned to edges entering v is the same as the product of elements assigned to the edges leaving v. Let $\mathsf{sflo}(G)$ be the number of S-flows. This number is independent of the orientation.

The choice $\Gamma = \mathbb{Z}_2$ and $S = \mathbb{Z}_2 \setminus \{0\}$ gives the special case above (incidence function of eulerian graphs). If $\Gamma = S = \mathbb{Z}_2$, then $\mathsf{sflo}(G)$ is the number of eulerian subgraphs of G. Perhaps the most interesting special case is when $|\Gamma| = t$ and $S = \Gamma \setminus \{0\}$, which gives the number of nowhere zero t-flows.

Surprisingly, this parameter (which is an edge coloring model) can be described as a homomorphism function. Let Γ^* be the character group of Γ. Let H be the complete directed graph (with all loops) on Γ^*. Let $\alpha_\chi := 1/|\Gamma|$ for each $\chi \in \Gamma^*$, and let

$$\beta_{\chi\chi'} := \sum_{s \in S} \chi^{-1}(s)\chi'(s),$$

for $\chi, \chi' \in \Gamma^*$. Using arguments related to duality transformations of models in statistical physics (see e.g. [KW82] and references therein) one can show [FLS04] that this weighted graph H represents sflo in the sense that $\mathsf{sflo}(\cdot) = \hom(\cdot, H)$. The condition on S that it is closed under inversion can be dropped if we use homomorphisms of directed graphs (Section 3.5).

In statistical physics, negative – and more generally, complex – nodeweights correspond to complex magnetic fields, which arise in the well-known Lee-Yang theory of phase transitions [BBCKK04, LY52, YL52]. The next example

[Sze05] shows that it is also interesting in our context to extend the definition of weighted graphs by allowing negative nodeweights (a direction we will not pursue here except for this example).

Example 3.14 (Matchings revisited). We have seen that the number $\mathsf{pmatch}(G)$ of perfect matchings has exponential rank connectivity but is not reflection positive, and hence it is not a homomorphism function. However, consider the following weighted graph H_x: We take a looped complete graph on two nodes u and v, and define

$$\alpha(u) = \frac{1}{x}, \quad \alpha(v) = -\frac{1}{x},$$

and

$$\beta(uu) = x + 1, \quad \beta(uv) = \beta(vv) = 1.$$

Then the following surprising fact holds:

$$\lim_{x \to 0} \hom(G, H_x) = \mathsf{pmatch}(G)$$

for every graph G.

4 Convergence and Limit

4.1 Quasi-random Graphs

Quasirandom (also called pseudorandom) graphs were introduced by Thomason [Tho87] and Chung, Graham and Wilson [CGW89]. These graphs have many properties that true random graphs have.

A sequence $(G_n : n = 1, 2, \dots)$ of graphs is called *quasirandom with density p* (where $0 < p < 1$), if for every simple finite graph F,

$$t(F, G_n) = (1 + o(1))p^{|E(F)|}. \tag{25}$$

(this is the asymptotic number of labeled copies of F in a random graph with edge probability p). The definition is usually formulated in terms of the number of injections (labeled copies) of F into G, but the two differ only in lower order terms, which are swallowed by the $o(1)$ in the definition.

It turns out that (25) implies many other properties that are familiar from the theory of random graphs; for example, almost all degrees are about pn, almost all codegrees are about p^2n etc. Many of these properties characterize quasirandom graphs, and so these provide many equivalent ways to define a quasirandom sequence [CGW89, Tho87]. Quasirandomness is closely related to Szemerédi's lemma [SS03, SS97]. One of the most surprising facts proved in [CGW89] is that it is enough to require the condition about the number of copies of F for just two graphs, namely K_2 (which just defines the edge density p) and the 4-cycle C_4. This fact can be stated and proved in a simpler way using the "limit set" $\mathcal{T}_0$ defined in Section 2.1:

Theorem 4.1. *If $t \in \mathcal{T}_0$ satisfies $t(C_4) = t(K_2)^4$, then for every simple graph H,*

$$t(H) = t(K_2)^{|E(H)|}.$$

In other words, $t(H)$ is the expected profile of a random graph $G(n,p)$ with $p = t(K_2)$; it is also the profile of the weighted graph consisting of a single node and a loop with weight p (the weight of the node does not matter).

To illustrate the power of reflection positivity, we give a proof of this theorem (the proof goes along the lines of the original, just the details are simpler).

Proof. Let $p = t(K_2)$. We first prove the conclusion for stars $K_{1,j}$:

$$t(K_{1,j}) = p^j. \tag{26}$$

Starting with $t(K_{1,2})$, let us first consider the connection matrix $M(t,1)$ and its 2×2 submatrix formed by the rows and columns corresponding to the graph K_1 and K_2. Positive semidefiniteness of this matrix gives $t(K_{1,2}) \geq t(K_2)^2 = p^2$. On the other hand $t(K_{1,2}) \leq \sqrt{t(C_4)} = p^2$ by positive semidefiniteness of the 2×2 submatrix of $M(t,2)$ indexed by $K_{1,2}$ (with its endpoints labeled) and $\overline{K}_2$, the empty graph on two nodes. The above two inequalities give $t(K_{1,2}) = p^2$, which proves (26) for $j = 2$. To prove the identity for $j > 2$, we again consider the connection matrix $M(t,1)$. By the identity we just established, its 2×2 submatrix formed by the rows and columns corresponding to the graph K_1 and K_2 has 0 determinant. By positive semidefiniteness, the corresponding two rows of the whole connection matrix $M(t,1)$ are proportional. But this means that

$$t(K_{1,j+1}) = pt(K_{1,j})$$

for every j, from which (26) follows by induction.

Next we show that for all complete bipartite graphs $K_{2,j}$:

$$t(K_{2,j}) = p^{2j}. \tag{27}$$

Since $t(K_{2,2}) = t(C_4) = p^4$ by assumption, this is true for $j = 2$. For general j, it follows just like (26) from the positive semidefiniteness of the matrix $M(t,2)$.

Now we prove the general case by a similar induction. Let us view the graph H as glued together from a star $K_{1,d}$ and a graph F on one fewer nodes, along the set T of the leaves of the star, and suppose that we know the assertion for F. Consider the matrix $M(t,d)$ and its 2×2 submatrix formed by the rows and columns indexed by $\overline{K}_d$ (the graph on d labeled nodes with no edge) and $K_{1,d}$ (with the leaves labeled). By (26) and (27), this submatrix is singular, and hence these two rows of the whole matrix are proportional, the second row is p^d times the first. But the graphs F and H define two elements of these rows above each other, so

$$t(H) = p^d t(F) = p^d p^{|E(F)|} = p^{|E(H)|},$$

which proves the theorem. □

4.2 Convergent Sequences

Let (G_n) be a sequence of unweighted simple graphs, and assume again that $|V(G_n)| = n$. We say that this sequence is *convergent*, if the sequence $t(F, G_n)$ has a limit for every simple graph F. Note that it would be enough to assume this for connected graphs F.

In the definition, we could replace the homomorphism function by the number of embeddings (injective homomorphisms), with appropriate normalization. Indeed, the difference between the number of homomorphisms and embeddings is the number of non-injective homomorphisms, which is of lower order, so it tends to 0 when divided by $n^{|V(F)|}$.

We could also replace the homomorphism function by the number of embeddings as induced subgraphs. Indeed, the number of embeddings can be obtained by summing the numbers of induced embeddings over all supergraphs (on the same set of nodes). Conversely, the number of induced embeddings can be expressed in terms of the numbers of embeddings of supergraphs by inclusion-exclusion.

Example 4.2. Let $\mathbf{G}(n,p)$ be a random graph on n nodes with edge-density p; the sequence $(\mathbf{G}(n,p), n = 1, 2, \dots)$ is convergent with probability 1. The limiting simple graph parameter is given by $t(F) = p^{|E(F)|}$.

By definition, every quasirandom graph sequence with density p is also convergent, and the homomorphism densities into it tend to the same value.

4.3 Finite Limits, a.k.a. Generalized Quasirandom Graphs

A *generalized random graph* $\mathbf{G}(n; H)$ is defined by the number n of its nodes and by a weighted "model" graph H. We assume that $V(H) = [q]$ and set $\alpha_i = \alpha_i(H)$ $(i = 1, \dots, q)$ and $\beta_{ij} = \beta_{ij}(H)$ $(i, j = 1, \dots, q)$. We partition $[n]$ into q classes $V_1, \dots, V_q$, by putting each $u \in [n]$ into V_i with probability α_i and connect each pair $u \in V_i$ and $v \in V_j$ with probability β_{ij} (all these decision are made independently).

A *generalized quasirandom graph sequence* (G_n) *with model graph* H (or briefly H*-quasirandom sequence*) is defined by the property that for every fixed finite graph F,

$$t(F, G_n) \longrightarrow t(F, H) \qquad (n \to \infty).$$

In other words, the number of homomorphisms of F into G_n is approximately the same as the expected number of homomorphisms of F into a generalized random graph $\mathbf{G}(N, H)$ on $N = |V(G_n)|$ nodes.

This definition suggests that we should consider the graph H as the "limit" of the H-quasirandom sequence. The definition of a quasirandom sequence of graphs (with edge-density p) is equivalent to saying that the sequence converges to $K_1(p)$. (Warning: not every convergent sequence will have a limit of this form!)

In view of the theory of quasirandom graphs, we can ask the following two basic questions concerning generalized quasirandom graphs:

(a) Is it enough to require the condition concerning the number of copies of F for a finite set of graphs F_i (depending on α and β)?
(b) Is the structure of a generalized quasirandom graph G_n similar to a generalized random graph?

To be more precise, we want that the nodes of G_n can be partitioned into q classes $U_1, \ldots, U_q$ of sizes $\alpha_1 n, \ldots, \alpha_q n$ so that the graph spanned by U_i is quasirandom with density β_{ii}, and the bipartite graph formed by the edges between U_i and U_j is quasirandom with density β_{ij}.

The answer to the first two questions is in the affirmative. More precisely, the following theorems hold.

Theorem 4.3 ([LSó06⁺]). *Let H be a weighted graph with $V(H) = [q]$, nodeweights $(\alpha_i : i = 1, \ldots, q)$ and edgeweights $(\beta_{ij} : i, j = 1, \ldots, q)$. Let $(G_n, n = 1, 2, \ldots)$ be an H-quasirandom sequence of unweighted simple graphs. Then for every n there exists a partition $V(G_n) = \{U_1, \ldots, U_q)$ such that*

(a) $\dfrac{|U_i|}{|V(G_n)|} \to \alpha_i \qquad (i = 1, \ldots, q)$,
(b) *the subgraph of G_n induced by U_i is a quasirandom graph sequence with edge density β_{ii}, and*
(c) *the bipartite subgraph between U_i and U_j is a quasirandom bipartite graph sequence with density β_{ij}.*

Theorem 4.4 ([LSó06⁺]). *Let H be a weighted graph with $V(H) = [q]$. A sequence $(G_n, n = 1, 2, \ldots)$ is H-quasirandom if and only if*

$$t(F, G_n) \longrightarrow t(F, H) \qquad (t \to \infty)$$

for every graph F with at most $(10q)^q$ nodes.

4.4 The General Limit Object

Limits as Reflection Positive Parameters

We now turn to describing limits of general convergent graph sequences.

Let $\widetilde{\mathcal{T}}_0$ be the set of homomorphism density functions $t(\cdot, G)$ defined on simple (unweighted) graphs, where G is any simple unweighted target graph. Let $\mathcal{T}_0$ be the set of all graph parameters that are pointwise limits of graph parameters in $\widetilde{\mathcal{T}}_0$ (i.e., its closure in the product topology on $\mathbb{R}^{\mathcal{F}}$). It is not hard to see that $\mathcal{T}_0$ would not change if we allowed weighted target graphs with edge weights between 0 and 1.

The characterization of homomorphism functions (Theorem 3.6) extends to the limit, at least for simple graphs:

Theorem 4.5 ([LSz04]). *A simple graph parameter f is in $\mathcal{T}_0$ if and only f is normalized, multiplicative and reflection positive.*

Limits as Measurable Functions

Graph parameters in the set $\mathcal{T}_0$ can be represented as homomorphism functions into measurable functions [LSz04]. Let $\mathcal{W}$ denote the set of all bounded measurable functions $W: [0,1]^2 \to \mathbb{R}$ such that $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$. We also introduce the set

$$\mathcal{W}_0 = \{W \in \mathcal{W} : 0 \le W \le 1\}.$$

Theorem 4.6 ([LSz04]). *A simple graph parameter f is in $\mathcal{T}_0$ if and only if there is a function $W \in \mathcal{W}_0$ such that $f = t(\cdot, W)$.*

This function W is not unique: for example, $W(1-x, 1-y)$ will define the same graph parameter. More generally, if $\phi: [0,1] \to [0,1]$ is a measure preserving map (not necessarily bijective), then

$$W^{\phi}(x,y) = W(\phi(x), \phi(y))$$

defines the same parameter. The following theorem says that this is all: Let us call functions $W_1, W_2 \in \mathcal{W}$ *equal up to measure preserving transformation* if there is a third function $W \in \mathcal{W}$ and measure preserving maps $\phi_1, \phi_2 : [0,1] \to [0,1]$ such that $W_i = W^{\phi_i}$.

Theorem 4.7 ([BCL06$^+$]). *Two functions $W_1, W_2 \in \mathcal{W}$ define the same simple graph parameter if and only if they are equal up to measure preserving transformation.*

Limits as Distributions over Finite and Countable Graphs

Given any function $W \in \mathcal{W}_0$ and an integer $n > 0$, we can generate a random graph $\mathbf{G}(n, W)$, called a *W-random graph*, on node set $[n]$ as follows. We generate n independent numbers $X_1, \dots, X_n$ from the uniform distribution on $[0,1]$, and then connect nodes i and j with probability $W(X_i, X_j)$.

As a special case, if W is the identically p function, we get "ordinary" random graphs $\mathbf{G}(n,p)$. More generally, if $W = W_H$ for a (finite) weighted graph H, then $\mathbf{G}(n, W_H)$ is a quasirandom graph with model H.

Theorem 4.8 ([LSz04]). *With probability 1, the graph sequence $\mathbf{G}(n, W)$ is convergent, and its limit is the function W.*

Let us define a *random graph model* as a distribution $\mathbf{G}_n$ on simple graphs on $[n]$, for every $n \in \mathbb{Z}_+$. The random graph model $\mathbf{G}(n, W)$ defined above has the following three obvious properties:

(i) The distribution of $\mathbf{G}_n$ is invariant under relabeling nodes;
(ii) If we delete node n from $\mathbf{G}_n$, the distribution of the resulting graph is the same as the distribution of $\mathbf{G}_{n-1}$;

(iii) for every $1 < k < n$, the subgraphs of $\mathbf{G}$ induced by $[k]$ and $\{k+1, \ldots, n\}$ are independent (as random variables).

It turns out that these three properties characterize the model $\mathbf{G}(n, W)$:

Theorem 4.9 ([LSz04]). *A random graph model is of the form* $\mathbf{G}(n, W)$ *for some function* $W \in \mathcal{W}_0$ *if and only if it satisfies conditions* (i), (ii) *and* (iii). *Furthermore, two functions* $W_1, W_2 \in \mathcal{W}_0$ *define the same random graph model if and only if they are equal up to measure preserving transformation.*

We have seen that the limit of ordinary random graphs $\mathbf{G}(n, 1/2)$ is the function $W \equiv 1/2$. It is, however, quite natural to think that the limit of ordinary random graphs should be the Rado graph (the countable random graph). It turns out that this is also true in the following sense: For every $W \in \mathcal{W}_0$, we can define a countable random graph $\mathbf{G}(\omega, W)$ on $\mathbb{Z}_+$, by choosing an infinite sequence $(X_0, X_1, \ldots)$ of independent uniform samples from $[0, 1]$, and connecting i and j with probability $W(X_i, X_j)$. A *countable random graph model* is a probability distribution on graphs on $\mathbb{Z}_+$ (with the σ-algebra generated by cylinders consisting of all graph containing a given edge).

Theorem 4.10 ([LSz06⁺]). *A countable graph model is of the form* $\mathbf{G}(\omega, W)$ *if and only if it satisfies* (i) *and* (iii) *above. Furthermore, the countable graph model* $\mathbf{G}(\omega, W)$ *determines* W *up to measure preserving transformation.*

Thus it is justified to say that with probability 1, $\mathbf{G}(n, 1/2)$ converges to the Rado graph $\mathbf{G}(\omega, 1/2)$. A word of caution is warranted here: as an unlabeled graph, $\mathbf{G}(\omega, 1/2)$ is isomorphic to $\mathbf{G}(n, 1/3)$ with probability 1. So viewing the Rado graph as an unlabeled graph would not contain enough information to characterize the limit; we have to view it as a probability distribution over graphs on a fixed countable set of nodes.

Examples

Example 4.11. Consider the *half-graphs* $H_{n,n}$: they are bipartite graphs on $2n$ nodes $\{1, \ldots, n, 1', \ldots, n'\}$, where i is connected to j' if and only if $i \leq j'$. It is easy to see that this sequence is convergent. Indeed, let F be a simple graph with k nodes; we show that the limit of $t(F, H_{n,n})$ exists. We may assume that F is connected. If F is non-bipartite, then $t(F, H_{n,n}) = 0$ for all n, so suppose that F is bipartite; let $V(F) = V_1 \cup V_2$ be its (unique) bipartition. Then every homomorphism of F into H preserves the 2-coloring, and so the homomorphisms split into two classes: those that map V_1 into $\{1, \ldots, n\}$ and those that map it into $\{1', \ldots, n'\}$. By the symmetry of the half-graphs, these two classes have the same cardinality.

Now F defines a partial order P on $V(F)$, where $u \leq v$ if and only if $u = v$ or $u \in V_1$, $v \in V_2$, and $uv \in E$. With respect to this partial order, $\frac{1}{2}\hom(F, H_{n,n})$ is just the number of order-preserving maps from $V(F)$ to the chain $\{1, \ldots, n\}$, and so

$$2^{k-1}t(F,H_{n,n}) = 2^{k-1} \cdot \frac{\hom(F,H_{n,n})}{(2n)^k} = \frac{\frac{1}{2}\hom(F,H_{n,n})}{n^k}$$

is the probability that a random map of $V(F)$ into $\{1,\dots,n\}$ is order-preserving. As $n \to \infty$, the fraction of non-injective maps tends to 0, and hence it is easy to see that $2^{k-1}t(F,H_{n,n})$ tends to a number $2^{k-1}t(F)$, which is the probability that a random ordering of $V(F)$ is compatible with P. In other words, $k!2^{k-1}t(F)$ is the number of linear extensions of P.

However, the half-graphs do not converge to any finite weighted graph. To see this, let S_k denote the star on k nodes, and consider the (infinite) matrix M defined $M_{k,l} = t(S_{k+l-1})$. If $t(F) = t(F, G_0)$ for some finite weighted graph G_0, then it follows from the characterization of homomorphism functions in [FLS04] that this matrix has rank at most $|V(G_0)|$; on the other hand, it is easy to compute that

$$M_{k,l} = \frac{1}{2^{k+l-2}(k+l-1)},$$

and this matrix (up to row and column scaling, the *Hilbert matrix*) has infinite rank (see e.g [Choi83]).

One can say, however, that in the limit, we are considering order-preserving maps of the poset P into the interval $[0, 1]$; equivalently, adjacency-preserving maps of F into the infinite graph with node set $[0, 1]$ and edge-set $\{xy : x \le 1/2, y > 1/2, x \le y - 1/2\}$. More precisely, the limit is given by the function

$$W(x,y) = \begin{cases} 1, \text{ if } x \ge y + \frac{1}{2} \text{ or } y \ge x + \frac{1}{2}, \\ 0, \text{ otherwise.} \end{cases}$$

Example 4.12 (Preferential attachment graphs). Preferential attachment as a paradigm of generating power law distributions goes back to Yule [Yule24] in 1923. In the context of networks and graph theory, the idea is usually credited to Barabasi and Albert [BA99], even though it can be already found in [Pri76]. For the simplest case of preferential attachment trees, the rigorous mathematical analysis of these models was first carried out in [MSS93], while the rigorous analysis of the more general model of Barabasi and Albert was first carried out in [BRST01]. For more general models of undirected preferential attachment graphs see [CF03], and for models with directed edges, see [BBCR03]. All these model are sparse models with bounded average degree.

Here we define a *preferential attachment graph* $\mathrm{PAG}(n, m)$ as the random graph with n nodes and m edges obtained by the following procedure. Fix a set of n nodes, and let $v_1 \dots v_n$ be any ordering of the nodes. We extend this sequence one by one by picking an element of the current sequence randomly and uniformly, and append a copy of it at the end. We repeat this until $2m$ further elements have been added. So we get a sequence $v_1 \dots v_n v_{n+1} \dots v_{n+2m}$.

Now we construct G by connecting nodes v_{n+2k-1} and v_{n+2k} for $k = 1, 2, \dots, m$, to get $G(n, m)$. (Note that G may have multiple edges and loops, which we have to live with for the time being).

Another way of describing this construction is to view it as adding edges one by one, where the probability of adding an edge connecting u and v is proportional to the product of the "degrees". To be more precise, the probability that the $(k+1)$-st edge connects u and v is

$$\begin{cases} \dfrac{2(d(u)+1)(d(v)+1)}{(n+2k)(n+2k+1)} & \text{if } u \neq v, \\ \dfrac{(d(u)+1)(d(u)+2)}{(n+2k)(n+2k+1)} & \text{if } u = v, \end{cases}$$

where $d(u)$ is the current degree of the node (adding 1 to the degree is needed to start the procedure at all; adding 2 to the second factor in the case when $u = v$ makes everything come out nicer).

It can be shown that the limit of preferential attachment graphs $\mathrm{PAG}(n, cn^2)$, with probability 1, is the function $W_c(x,y) = c(\log x)(\log y)$. It is interesting to note that the graphs $\mathbf{G}(n, W_c)$ form another (different) sequence of random graphs tending to the same limit W_c with probability 1.

5 The Metric Space of Graphs

5.1 Distances of Graphs

In this section, we assume that the graph G (which we probe by F from the left and by H from the right) is dense, i.e., the number of edges of G is $\Omega(n^2)$, where n is the number of nodes (the results are valid but mostly vacuous for sparse graphs). We'll discuss analogous questions for sparse graphs (specifically, for graphs with bounded degree) in Section 8.

Matrix Norms

For an $n \times n$ matrix A, the *rectangle norm* (also called the *cut norm*) is defined as

$$\|A\|_\square = \max_{u,v \in \{0,1\}^n} |u^\top A v|. \tag{28}$$

This norm is closely related to $\ell_\infty \to \ell_1$ norm, which can be defined by

$$\|A\|_{\infty\to 1} = \max_{u,v \in [-1,1]^n} |u^\top A v| = \max_{u,v \in \{-1,1\}^n} u^\top A v; \tag{29}$$

in fact,

$$\frac{1}{4}\|A\|_{\infty\to 1} \leq \|A\|_\square \leq \|A\|_{\infty\to 1}. \tag{30}$$

For symmetric matrices A, the norm

$$\|A\|'_\square = \max_{u \in \{0,1\}^n} |u^\top A u| \tag{31}$$

is simpler to define and is also equivalent to the rectangle norm:

$$\|A\|'_\square \le \|A\|_\square \le 2\|A\|'_\square. \tag{32}$$

A constant factor approximation of the rectangle norm can be computed in polynomial time, using semidefinite optimization and Grothendieck's inequality in functional analysis (see [AN06+]).

Labeled Graphs on the Same Set of Nodes

Let G and G' be two graphs on the same set of n nodes. We want to define a notion of distance between them that reflects structural similarity. A first attempt is to define

$$d_1(G, G') = \frac{1}{n^2}|E(G)\triangle E(G')|.$$

(Here the division by n^2 is just a convenience, so that the distance of two graphs is always between 0 and 1.) However, this notion is too restrictive: For example, the distance of two random graphs with the same density is of constant order (with large probability), even though two random graphs are structurally very similar.

For our purposes, the following distance function will be more useful. For a graph G and sets $S, T \subseteq V(G)$, let $e_G(S,T)$ denote the number of edges in G with one endnode in S and the other in T (the endnodes may also belong to $S \cap T$; so $e_G(S,S)$ is twice the number of edges spanned by S). We define:

$$d_\square(G, G') = \frac{1}{n^2} \max_{S,T \subseteq V(G)} |e_G(S,T) - e_{G'}(S,T)|.$$

Note that we are dividing by n^2 and not by $|S| \times |T|$, so the contribution of a pair S, T is at most $|T| \times |S|/n^2$. Thus small sets of size $o(n)$ play no role when measuring the distance. In terms of the adjacency matrices A and A' of G and G', respectively, this can be expressed as

$$d_\square(G, G') = \frac{1}{n^2}\|A - A'\|_\square.$$

Note that the definition can be extended to the case when G and G' have edgeweights. Furthermore, by (30) and (32), we could replace the $\|.\|_\square$ norm in the definition by one of the other matrix norms defined above without distorting the distance by more than a constant factor.

We need to extend this notion to weighted graphs on the same set of nodes. Let G and G' be weighted graphs with $V(G) = V(G')$. We assume that G and G' both have total nodeweight 1, but the weights of individual nodes in G and G' may be different. Then we define

$$d_\square(G,G') = \sum_i |\alpha_i(G) - \alpha_i(G')| \qquad (33)$$
$$+ \max_{S,T\subseteq V(G)} \Big| \sum_{\substack{i\in S\\ j\in T}} \big(\alpha_i(G)\alpha_j(G)\beta_{ij}(G) - \alpha_i(G')\alpha_j(G')\beta_{ij}(G')\big) \Big|.$$

If the two graphs do not both have total nodeweights of one, then we simply define the distance $d_\square$ in terms of the corresponding "normalized" graphs, i.e., we replace $\alpha_i(G)$ with $\alpha_i(G)/\alpha(G)$, and similarly for G'.

Unlabeled Graphs with the Same Number of Nodes

Now assume that G and G' are unlabeled unweighted graphs on n nodes. It is natural to define

$$\widehat{\delta}_\square(G,G') = \min_{\tilde{G},\tilde{G}'} d_\square(\tilde{G},\tilde{G}'), \qquad (34)$$

where $\tilde{G}$ and $\tilde{G}'$ range over all labelings of G and G' by $1,\ldots,n$, respectively (of course, we could fix the labeling of one of the graphs).

Consider any labeling that attains the minimum in the definition of $\widehat{\delta}_\square$, and identify the nodes of G and G' with the same label. In this case, we say that G and G' are *optimally overlaid.*

Unlabeled Graphs with Different Number of Nodes

To define the distance (in any of the above senses) of two unlabeled graphs with different number of nodes, say G with n nodes and G' with n' nodes, a first idea is to blow up each node of G into n' nodes, and each node of G' into n nodes, so that both graphs now will have nn' nodes. An improved version of this idea is to match up the nodes "fractionally". This also allows us to extend the notion of distance to weighted graphs.

Let G and G' be weighted graphs with (say) $V(G) = [n]$, $V(G') = [n']$, and assume that the sum of nodeweights is 1 (just scale the nodeweights of each graph). Let X be a nonnegative $n \times n'$ matrix such that

$$\sum_{u=1}^{n'} X_{iu} = \alpha_i(G)$$

and

$$\sum_{i=1}^{n} X_{iu} = \alpha_u(G').$$

We think of X_{iu} as the portion of node i that is mapped onto node u. We call such a matrix X a *fractional overlay* of G and G'. Let $\mathcal{X}(G,G')$ denote the set of all fractional overlays. Note that for every $X \in \mathcal{X}(G,G')$,

$$\sum_{i=1}^{n}\sum_{u=1}^{n'} X_{iu} = \sum_{i=1}^{n} \alpha_i(G) = \sum_{u=1}^{n'} \alpha_u(G') = 1.$$

(If we view α_G and $\alpha_{G'}$ as probability distributions, then every $X \in \mathcal{X}(G,G')$ is a *coupling* of these distributions.)

For each fractional overlay, we construct the following two weighted graphs. The nodes of $G[X]$ are all pairs (i,u) where $1 \le i \le n$ and $1 \le u \le n'$. The weight of the node (i,u) is X_{iu}, and the weight of the edge $((i,u),(j,v))$ is β_{ij}. The other graph $G'[X^\mathsf{T}]$ is defined similarly, except that the roles of i and u are interchanged. Now the node sets of $G[X]$ and $G'[X^\mathsf{T}]$ are labeled by the same set of pairs (i,u), so their distances are well defined.

Thus we can define the distance of two weighted unlabeled graphs G and G' (with total nodeweight 1):

$$\delta_\square(G,G') = \min_{X \in \mathcal{X}(G,G')} d_\square(G[X], G'[X^\mathsf{T}]).$$

We can express this distance in terms of the original graphs G and G' by the following formula:

$$\delta_\square(G,G') = \min_{X \in \mathcal{X}(G,G')} \max_{S,T \subseteq V \times V'} \Big| \sum_{\substack{(i,u)\in S\\(j,v)\in T}} X_{iu}X_{jv}\big(\beta_{ij}(G) - \beta_{uv}(G')\big)\Big|. \quad (35)$$

Of course, this definition also applies if G and G' have the same number of nodes; however, it may give a different value than (34). It is proved in [BCLSV06$^+$] that there is a constant $c > 0$ such that

$$\delta_\square(G,G') \le \widehat{\delta}_\square(G,G') \le c\,\delta_\square(G,G')^{1/4} \quad (36)$$

(the lower bound is trivial; we do not have an example showing that the exponent 1/4 is needed in the upper bound). While the definition of $\widehat{\delta}_\square$ is more straightforward, the distance $\delta_\square$ will be easier to work with, and we will use mostly the latter distance.

A very special weighted graph is $K_1(p)$: a single node with a loop with weight p. For the random graph $\mathbf{G} = \mathbf{G}(n,p)$, we have a.s.

$$\delta_\square(\mathbf{G}, K_1(p)) \longrightarrow 0 \qquad (n \to \infty).$$

Let (G_n) be a sequence of simple graphs. It follows by standard results on quasirandom graphs that

Proposition 5.1 ([CGW89]). *A sequence (G_n) of graphs is quasirandom if and only if $\delta_\square(G_n, K_1(p)) \to 0$ as $n \to \infty$.*

The following result connects this distance to homomorphism functions.

Lemma 5.2 ([LSz04]). *For any three simple graphs F, G and G'*

$$|t(F,G) - t(F,G')| \le |E(F)| \cdot \delta_\square(G,G').$$

5.2 Szemerédi's Lemma

For a graph $G = (V, E)$ and two subsets $U, W \subset V$ we define the "irregularity" of the pair U, W as the quantity

$$\mathrm{irreg}_G(U, V) = \max_{X \subseteq U, Y \subseteq V} \big| e_G(X, Y) - d|X| \cdot |Y| \big|,$$

where d is the density $d = e_G(U, W)/(|U| \cdot |W|)$. Let $\mathrm{twr}(\varepsilon)$ denote the $\lceil 1/\varepsilon^2 \rceil$ times iterated exponential function (the "tower"). With this notation, we can state one version of the Regularity Lemma:

Lemma 5.3 (Szemerédi Regularity Lemma). *For every $\varepsilon > 0$ and every graph $G = (V, E)$ there is a partition $\mathcal{P}$ of V into $k \leq \mathrm{twr}(\varepsilon)$ classes $V_1, \dots, V_k$ such that*

$$\sum_{1 \leq i < j \leq k} \mathrm{irreg}_G(V_i, V_j) \leq \varepsilon |V|^2.$$

(While this form is perhaps easiest to state and prove, there are equivalent forms that are more suited for applications. We refer to [KS96] for a survey, to [SS91] for connections with quasirandom graphs, and to [LSz05c] for analytic aspects of the Regularity Lemma.)

The appearance of the tower function in the Lemma forbids practical applications (and unfortunately this bound on the number of parts is not far from best possible, as it was shown by Gowers [Gow97]). A more reasonable threshold was proved by Frieze and Kannan [FK99], at the cost of using a weaker measure of irregularity.

Given a graph $G = (V, E)$ and a partition $\mathcal{P} = \{V_1, \dots, V_k\}$ of V, we define a weighted graph $G_{\mathcal{P}}$ on V by setting $\alpha_u(G_{\mathcal{P}}) = 1$ and $\beta_{uv}(G_{\mathcal{P}}) = d_{i(u)i(v)}$, where $i(u)$ is the index of the set V_i containing u and $d_{ij} = e_G(V_i, V_j)/(|V_i||V_j|)$. The edgeweight matrix of $G_{\mathcal{P}}$ is thus obtained from the adjacency matrix of G by replacing each entry in the block $V_i \times V_j$ by the average over the block. In this notation, the result of Frieze and Kannan [FK99] can be formulated as follows:

Lemma 5.4 (Weak Regularity Lemma). *For every $\varepsilon > 0$ and every graph $G = (V, E)$, there exists a partition $\mathcal{P}$ of V into $k \leq 2^{2/\varepsilon^2}$ classes such that $d_\square(G, G_{\mathcal{P}}) \leq \varepsilon$.*

The bound on the number of partition classes is still rather large (exponential), but at least not a tower. Frieze and Kannan show that the partition can be obtained as an "overlay" of only $1/\varepsilon^2$ sets, so it has a description that is polynomial in $1/\varepsilon$, which in some applications leads to polynomial time algorithms (see e.g. [AVKK03]).

The Weak Regularity Lemma immediately implies the following slight variant:

Lemma 5.5 ([BCLSV06+]). *For every* $\varepsilon > 0$ *and every graph* $G = (V, E)$*, there is a weighted graph* H *with at most* $\lceil 2^{1/\varepsilon^2} \rceil$ *nodes such that* $\delta_\square(G, H) \le \varepsilon$*.*

We note that other versions strengthen the conclusion (of course, at the cost of replacing the tower function by an even more huge value). Such a "super-strong" Regularity Lemma was proved and used by Alon and Shapira [AS06+]. It would be interesting to fit the original Regularity Lemma or one of its applications into this framework.

5.3 Sampling from a Graph

The following important fact connecting sampling and graph distance follows from a result in [AVKK03]; see also [BCLSV06+] for a simple proof of (a):

Theorem 5.6. *Let* G_1 *and* G_2 *be two graphs on the same set of nodes* V*, let* $\varepsilon = d_\square(G_1, G_2)$*,* $\delta > 0$*, and let* S *be a random* k*-subset of* V*.*

(a) *If* $k \ge \frac{300}{\varepsilon^2} \log(\frac{2}{\delta})$*, then with probability at least* $1 - \delta$*,*

$$2^{-10}\varepsilon \le d_\square(G_1[S], G_2[S]) \le 4\varepsilon^{1/4}.$$

(b) *If* $k \ge 10^{10} \log(2/\varepsilon)/(\varepsilon^4\delta^5)$*, then with probability at least* $1 - \delta$*,*

$$2^{-10}\varepsilon \le d_\square(G_1[S], G_2[S]) \le 10^7 \frac{\varepsilon}{\sqrt{\delta}}.$$

Using this bound and the (weak) Regularity Lemma, it is not hard to prove the following theorem, which is the key to several further results.

Theorem 5.7 ([BCLSV06+]). *Let* G *be a (possibly weighted) graph,* $\varepsilon > 0$ *and* $k \ge 2^{c_1/\varepsilon^8}$*. Let* S *be a random subset of* $V(G)$ *of size* k*. Then with probability at least* $1 - \varepsilon$*, we have* $\delta_\square(G, G[S]) < \varepsilon$*.*

Informally, if we take a sample of k points, and blow up each node of this subgraph into $|V(G)|/k$ twins, then the resulting graph can be overlayed with G so that the $d_\square$-distance will be small.

This theorem can be thought of as a strengthening of the (weak) Regularity Lemma in two directions. First, it says that the approximating weighted graph can be required to be unweighted; second, that it can be obtained just by drawing a random sample.

From this theorem, it is easy to deduce the following converse of Lemma 5.2:

Theorem 5.8 ([BCLSV06+]). *Let* G*,* G' *be simple graphs, and let* $\varepsilon > 0$*. Set* $k = \lceil 2^{c_2/\varepsilon^8} \rceil$*, and assume that for every simple graph* F *on at most* k *nodes, we have* $|t(F, G) - t(F, G')| < 2^{-2k^2}$*. Then* $\delta_\square(G, G') \le \varepsilon$*.*

These results allow us to characterize convergent graph sequences:

Theorem 5.9 ([BCLSV06⁺]). *A graph sequence is convergent if and only if it is Cauchy in the* $\delta_\square$ *metric.*

Let $\mathcal{F}$ denote the metric space of all finite, simple graphs with the $\delta_\square$ metric. It follows from the above that the completion $\mathcal{X}$ of $\mathcal{F}$ can be described as follows. Consider the space of all functions in $\mathcal{W}_0$, with the distance

$$d_\square(U,W) = \sup_{S,T\subseteq[0,1]} \Big| \int_{S\times T} W(x,y)dx\,dy \Big|.$$

Define a new metric by

$$\delta_\square(U,W) = \inf d_\square(U^\phi, W^\psi),$$

where ϕ and ψ range over all measure preserving maps $\phi,\psi:\ [0,1]\to[0,1]$. Then the elements of $\mathcal{X}$ can be obtained by identifying functions that are at distance 0, and the $\delta_\square$ metric between these classes extends the $\delta_\square$ metric on graphs.

The above results give various other descriptions of this completion: for example, $\mathcal{X}$ is isomorphic to $\mathcal{T}_0$ with the metric

$$\delta(t_1,t_2) = \sup_F \frac{1}{|E(F)|}|t_1(F) - t_2(F)| \qquad (t_1,t_2\in\mathcal{T}_0).$$

Szemerédi's Lemma can be used to show that $\mathcal{X}$ is compact.

5.4 Testing Huge Graphs

Imagine that we have a huge graph G; this graph is so large that we cannot describe it completely in any way. All we can do is sample a bounded number of nodes of G and look at the subgraph that is induced by them. What can we learn about G?

There are two related, but slightly different ways of asking this question.

Parameter Testing

Parameter testing is easier to state. We may want to determine some parameter of G; say what is the edge density? How large is the density of the maximum cut? Of course, we will not be able to determine the exact value of this parameter; the best we can hope for is that if we take a sufficiently large sample, we can find the approximate value of the parameter with large probability. To be precise, a graph parameter f is *testable*, if for every $\varepsilon > 0$ there is a positive integer k such that if G is a graph with at least k nodes and we select a set X of k independent uniform random nodes of G, then from the subgraph induced by them we can compute an estimate $\widetilde{f}(G[X])$ of f such that

$$\mathsf{P}(|f(G) - \widetilde{f}(G[X])| > \varepsilon) < \varepsilon.$$

It is an easy observation that we can always use $\widetilde{f}(G[X]) = f(G[X])$.

Using the notions of graph distance and convergence introduced above, we can give a number of characterizations of testable parameters.

Proposition 5.10 ([BCLSV06$^+$]). *A simple graph parameter is testable if and only if any of the following equivalent conditions holds.*

(a) *For every convergent graph sequence (G_n), the limit of $f(G_n)$ exists as $n \to \infty$ (continuity at infinity).*

(b) *For every $\varepsilon > 0$ there is an integer k_0 such that for every $k > k_0$ and every graph G on at least k nodes, a random set X of k nodes of G satisfies*

$$|f(G) - \mathsf{E}(f(G[X]))| < \varepsilon.$$

(c) *f is "essentially" uniformly continuous with respect to the $\delta_\square$ distance in the following sense: For every $\varepsilon > 0$ there is an $\varepsilon_0 > 0$ and a positive integer n_0 so that if G_1 and G_2 are two graphs with $|V(G_i)| \geq n_0$ and $\delta_\square(G_1, G_2) < \varepsilon_0$, then $|f(G_1) - f(G_2)| < \varepsilon$.*

(d) *There exists a functional $\widehat{f}(W)$ on $\mathcal{W}_0$ that is continuous in the rectangle norm, and extends f in the sense that $|\widehat{f}(W_G) - f(G)| \to 0$ if $|V(G)| \to \infty$.*

If we want to use (c) to prove that a certain invariant is testable, then the complicated definition of the $\delta_\square$ distance may cause a difficulty. So it is useful to show that (c) can be replaced by a weaker condition, which consists of three special cases of (c):

Supplement 5.11 ([BCLSV06$^+$]). *The following three conditions together are also equivalent to the testability:*

(c.1) *For every $\varepsilon > 0$ there is an $\varepsilon' > 0$ such that if G and G' are two simple graphs on the same node set and $d_\square(G, G') \leq \varepsilon'$ then $|f(G) - f(G')| < \varepsilon$.*

(c.2) *For every simple graph G, $f(G(m))$ has a limit as $m \to \infty$, where $G(m)$ denotes the graph obtained from G by replacing each node by m twins.*

(c.3) *$f(G \cup K_1) - f(G) \to 0$ if $|V(G)| \to \infty$.*

Some of the implications between conditions (a)–(d) in the Theorem are easy, some others follow from the general theory sketched above. To illustrate the use of this theorem, let us consider the density of the maximum cut:

$$f(G) = \max_{S \subseteq V(G)} \frac{e_G(S, V(G) \setminus S)}{\binom{n}{2}}.$$

This parameter is testable: this fact is nontrivial, and its first proof by Goldreich, Goldwasser and Ron [GGR98] was one of the first important results in Property Testing. Of the conditions above, (a) and (b) are more or less a reformulation of testability.

Condition (c), on the other hand, is easy to verify in this case. Let $\varepsilon > 0$, and let G_1 and G_2 be two graphs for which $\delta_\square(G_1, G_2) < \varepsilon$. Let us blow up the points of each graph so that the new graphs G_1' and G_2' have the same number N of points and they can be overlaid so that $d_\square(G_1', G_2') < \varepsilon$. For any subset $S \subseteq V(G_1') = V(G_2')$, we have

$$|e_{G_1'}(S, V(G) \setminus S) - e_{G_2'}(S, V(G) \setminus S)| < \varepsilon N^2,$$

and hence

$$|f(G_1') - f(G_2')| < \varepsilon.$$

To complete the proof, one must argue that $|f(G_i') - f(G_i)|$ is small, which is not hard (and is not given here).

Condition (d) can also be directly verified: we can extend the definition of a maximum cut to functions $W \in \mathcal{W}_0$ in a natural way:

$$\widehat{f}(W) = \sum_{S \subseteq [0,1]} \int_S \int_{[0,1] \setminus S} W(x, y)\, dx\, dy.$$

Then it is easy to check that this functional is continuous in the norm $\|.\|_\square$, and extends f.

Property Testing

Instead of estimating a numerical parameter, we may want to determine some property of G: Is G 3-colorable? Is it connected? Does it have a triangle? The answer will of course have some uncertainty. A precise definition was given by Goldreich, Goldwasser and Ron [GGR98], who also proved several fundamental results about this problem. There are in fact several ways to formalize this question. For this exposition, we take the following.

As for parameter testing, we specify an $\varepsilon > 0$ and want to find a positive integer k (depending on ε) with the following property. We select k independent uniform random nodes of G, and from the subgraph induced by them we compute a guess $X \in \{YES, NO\}$. Ideally, we want that if the graph does have the property, our guess should be YES with large probability, and if the graph does not have the property, then we should guess NO with large probability. But this is too much to ask. Suppose that we have two graphs that can be obtained from each other by changing a very tiny fraction of the edges, but one has the property, and the other does not. Then a sample induced subgraph from one graph will have almost the same distribution as a sample (of the same size) from the other, and so our guess for the two graphs will be almost the same.

A graph property $\mathcal{P}$ is *testable*, if our guess satisfies the following: if a graph has the property in a robust way so that changing at most εn^2 edges in any way it still has the property, we must guess YES with probability at least $1 - \varepsilon$; similarly, if changing at most εn^2 edges in any way the obtained graph does not have the property, then we must guess NO with probability at least $1 - \varepsilon$; in the grey area inbetween, we can guess arbitrarily. In other words, whatever we guess, we should be able to change at most εn^2 edges to make out guess right.

Remark 5.12. While changing a small number of edges is the most natural way to formalize that there is a "nearby" graph with the property, we have seen that the rectangular distance is often better behaved. One is tempted to define that a property is *weakly testable*, if for every $\varepsilon > 0$ there is a k such that for every graph G on at least k nodes we can make a guess based on a sample induced subgraph of size k such that with probability at least $1 - \varepsilon$, there is a graph G' such that $d_\square(G, G') < \varepsilon$ and our guess is right for G'. But this notion is not very interesting due to the fact that *every graph property is weakly testable.* This is an easy application of Theorem 5.7 above.

Among the many results on graph property testing, let us quote a surprisingly general recent result of Alon and Shapira [AS06+]. A graph property is called *hereditary*, if it is inherited by induced subgraphs.

Theorem 5.13 ([AS06+]). *Every hereditary graph property is testable.*

They in fact obtain a stronger result, which can be cast in the framework of parameter testing. For a hereditary graph property $\mathcal{P}$, define the *distance from the property* as $d(G, \mathcal{P}) = D(G, \mathcal{P})/|V(G)|^2$, where $D(G, \mathcal{P})$ is the minimum number of edges we need to change in G to obtain a graph with property $\mathcal{P}$. Alon and Shapira proved:

Theorem 5.14 ([AS06+]). *The distance from a hereditary graph property is testable.*

This theorem has a reasonably short proof using graph limits, see [LSz05c].

6 Partitions and Homomorphisms into Small Graphs

6.1 Ground State Energy

In Section 4 we defined convergence of a graph sequence G_n in terms of the homomorphism numbers from small graphs F to G_n ("convergence from the left"). But there are many applications where one wants to study homomorphisms from G_n into a small graph H. This naturally raises the question whether suitably normalized homomorphism numbers $\hom(G_n, H)$ converge if G_n is convergent from the left.

Consider a graph G on n nodes, and a softcore graph H on q nodes. (Recall that H is called softcore if all edge weights are strictly positive.) Then

$\log \hom(G,H)$ typically grows like the number of edges in G. For dense graphs, it therefore seems natural to consider the quantity

$$\frac{1}{n^2} \log \hom(G,H). \tag{37}$$

This quantity is closely related to a weighted maximum cut problem on G. Indeed, let $B = (B_{ij})_{1 \le i,j \le q}$ be a symmetric matrix with real entries. We then define the *ground state energy* of the "model" B on the graph G as

$$\mathcal{E}(G,B) = \frac{1}{n^2} \max_{\phi\colon V(G) \to [q]} \sum_{uv \in E(G)} B_{\phi(u)\phi(v)}. \tag{38}$$

For large n, the quantity defined in (37) is well approximated by the ground state energy.

Lemma 6.1. *Let G be an unweighted graph with n nodes, and let H be a softcore graph with $\alpha(H) = 1$. Let $\alpha = \min_i \alpha_H(i)$. Let B be the matrix of logarithms of the edgeweights of H. Then*

$$\mathcal{E}(G,B) - \frac{\log(1/\alpha)}{n} \le \frac{\log \hom(G,H)}{n^2} \le \mathcal{E}(G,B).$$

(Note that the upper bound on $\log \hom(G,H)/n^2$ does not depend on the nodeweights of H, and in the lower bound, only the error term does.)

Proof. Note that

$$\max_{\phi\colon V(G) \to V(H)} \hom_\phi(G,H) = e^{n^2 \mathcal{E}(G,B)}.$$

Thus we have

$$\hom(G,H) = \sum_\phi \alpha_\phi \hom_\phi(G,H) \le \sum_\phi \alpha_\phi e^{n^2 \mathcal{E}(G,B)} = e^{n^2 \mathcal{E}(G,B)},$$

and

$$\hom(G,H) \ge \max_\phi \alpha_\phi \hom_\phi(G,H) \ge \alpha^n e^{n^2 \mathcal{E}(G,B)}.$$

From these bounds the Lemma follows. □

Example 6.2. As a special case, consider the graph H consisting of two nodes of weight $1/2$, with loops of weight 1 at each, and connected by an edge of weight e (the base of the natural logarithm). Then $n^2 \mathcal{E}(G,H')$ is the size of the maximum cut in G, which we denote by $\mathrm{MAXCUT}(G)$. By Lemma 6.1,

$$\log \hom(G,H) - n \le \mathrm{MAXCUT}(G) \le \log \hom(G,H).$$

Since $\mathrm{MAXCUT}(G) \ge |E(G)|/2$, this gives a very good approximation of the maximum cut. More generally, for a fixed H, computing $\mathcal{E}(G,H)$ is a weighted multiway cut problem.

Our next theorem states that convergence from the left implies convergence of the ground state energies. Its proof uses the notion of fractional partitions, a notion we will need at several places in this section: a *fractional partition* of a set V into q classes (briefly, a fractional q-partition) is a q-tuple $\rho = (\rho_1, \ldots, \rho_q)$ of functions from V to $[0,1]$ such that for all $x \in V$, we have $\rho_1(x) + \ldots + \rho_q(x) = 1$. (Later in this section, we will apply this definition also to the case when $V = [0,1]$, when we tacitly assume that the functions ρ_i are measurable.) We will use the notation Pd_q for the set of probability distributions on $[q]$, i.e., the set of vectors $a = (a_1, \ldots, a_q)$ such that $a_i \geq 0$ and $\sum_i a_i = 1$, and the notation Sym_q for the set of $q \times q$ symmetric matrices.

Theorem 6.3 ([BCLSV06+]). *Let q be a positive integer, let $B \in \mathsf{Sym}_q$, and let G_n be a convergent sequence of simple graphs. Then $\mathcal{E}(G_n, B)$ is a convergent sequence.*

As a simple illustration of the usefulness of Proposition 5.10 and its Supplement 5.11, we sketch the proof. Indeed, let us consider the quantity

$$\mathcal{E}_\phi(G,B) = \frac{1}{n^2} \sum_{uv \in E(G)} B_{\phi(u)\phi(v)}$$

where ϕ is a map from $V(G)$ to $[q]$. Identifying these maps with partitions $\mathcal{P} = (V_1, \ldots, V_q)$ of $V(G)$ and using the definition of the $d_\square$ metric, we immediately see that

$$|\mathcal{E}_\phi(G,B) - \mathcal{E}_\phi(G',B)| \leq q^2 \|B\|_\infty d_\square(G,G')$$

whenever G and G' are simple graphs on the same set of nodes, which verifies the condition (c.1) of Supplement 5.11.

The condition (c.2) is not hard to verify either: Define the energy of a fractional q-partition ρ of $V(G)$ as

$$\mathcal{E}_\rho(G,B) = \frac{1}{n^2} \sum_{uv \in E(G)} \sum_{i,j \in [q]} \rho_i(u)\rho_j(v) B_{ij}. \tag{39}$$

Then $\mathcal{E}(G(k),B) = \max_\rho \mathcal{E}_\rho(G,B)$ where the maximum runs over all fractional partitions such that all $\rho_i(u)$'s are multiples of $1/k$. We claim that the maximum is attained for fractional partitions which are $\{0,1\}$ valued, so that $\mathcal{E}(G(k),B) = \mathcal{E}(G,B)$ for all k. It is clear from the above description that $\mathcal{E}(G(k),B) \geq \mathcal{E}(G,B)$, so the only thing we need to show is a matching upper bound on $\mathcal{E}(G(k),B)$. Consider a fractional partition ρ maximizing $\mathcal{E}_\rho(G,B)$, and a fixed node $u \in V(G)$. Then $\mathcal{E}_\rho(G,B)$ is a linear function of the vector $(\rho_1(u), \ldots, \rho_q(u))$, implying that the maximum over all these vectors is obtained at a vertex of the simplex Pd_q. Applying this procedure to all nodes $u \in V(G)$, this gives the desired inequality $\mathcal{E}(G(k),B) \leq \mathcal{E}(G,B)$, and hence the equality of $\mathcal{E}(G(k),B)$ and $\mathcal{E}(G,B)$ for all k. The condition (c.2) therefore holds trivially. The verification of condition (c.3) is even easier.

Let us finally note that the limiting ground state energy of a convergent sequence can be expressed explicitly. Indeed, for $W \in \mathcal{W}_0$ and a symmetric $q \times q$ matrix B, let

$$\mathcal{E}(W,B) = \max_{\rho} \frac{1}{2} \sum_{i,j=1}^{q} B_{ij} \int_{[0,1]^2} \rho_i(x)\rho_j(y)W(x,y)\,dx\,dy, \tag{40}$$

where the maximum runs over fractional q-partitions of $[0,1]$. Then we have the following theorem.

Theorem 6.4 ([BCLSV06$^+$]). *Let (G_n) be a convergent sequence, and let $W \in \mathcal{W}_0$ be its limit. Let H be a softcore graph, and let B be the matrix of logarithms of the edgeweights of H. Then*

$$\lim_{n\to\infty} \frac{\log \hom(G_n,H)}{|V(G_n)|^2} = \lim_{n\to\infty} \mathcal{E}(G_n,B) = \mathcal{E}(W,B).$$

6.2 Entropy and Free Energy

In addition to the ground state energy and the closely related quantity (37), we will consider the so-called "pressure" or "free energy" of the model H on G, defined as

$$\widehat{\mathcal{P}}(G,H) = \frac{1}{n} \log \hom\big(\frac{1}{n}G,H\big) \tag{41}$$

where n is the number of nodes in G and $\frac{1}{n}G$ is obtained from G by multiplying the edge weights of G by $\frac{1}{n}$.

To discuss the convergence of the free energy, we need the notion of the entropy of fractional partitions. Let $\rho = (\rho_1,\dots,\rho_q)$ be a fractional partition of a finite set V. Then the entropy of ρ is defined as

$$\mathcal{H}(\rho) = -\frac{1}{|V|} \sum_{u\in V} \sum_{i=1}^{q} \rho_i(u) \log \rho_i(u).$$

Theorem 6.5 ([BCLSV06$^+$]). *Let q be a positive integer, let H be a softcore weighted graph, and let (G_n) be a convergent sequence of simple graphs. Then $\widehat{\mathcal{P}}(G_n,H)$ is a convergent sequence.*

If G_n converges to a function $W \in \mathcal{W}_0$, the limiting free energy can again be expressed as an explicit function of W. To this end, let us define the entropy of a fractional q-partition ρ of $[0,1]$ as

$$\mathcal{H}(\rho) = -\int_0^1 \sum_{i=1}^{q} \rho_i(x) \log \rho_i(x)\,dx,$$

and the entropy of the H colorings of the limit graph W as

$$\mathcal{P}(W,H) = \max_\rho\Big(\mathcal{H}(\rho) + \sum_i h_i \int_{[0,1]} \rho_i(x)\,dx + \frac{1}{2}\sum_{i,j} B_{ij} \int_{[0,1]^2} W(x,y)\rho_i(x)\rho_j(y)\,dx\,dy\Big), \qquad (42)$$

where the maximum goes over all fractional q-partitions of $[0,1]$, $h_i = \log\alpha_i(H)$ and $B_{ij} = \log\beta_{ij}(H)$.

Theorem 6.6 ([BCLSV06⁺]). *Let (G_n) be a convergent sequence of simple graphs, and let $W \in \mathcal{W}_0$ be its limit. Let H be a weighted softcore graph. Then*

$$\lim_{n\to\infty} \frac{1}{n}\log \hom(\frac{1}{n}G_n, H) = \mathcal{P}(W,H).$$

This theorem can be proved using Proposition 5.10 and its Supplement 5.11, in a way similar to (but more involved than) the proof of Theorem 6.3.

6.3 Factor Graphs

Let G be a weighted graph and let $\mathcal{P} = (V_1,\dots,V_q)$ be a partition of $V(G)$. The *factor graph* (or briefly *factor*) $H(G,\mathcal{P})$ is the weighted graph on $[q]$ with nodeweights

$$\alpha_i(H(G,\mathcal{P})) = \frac{\alpha(G[V_i])}{\alpha(G)} = \frac{\sum_{u\in V_i}\alpha_u(G)}{\alpha(G))},$$

and edges weights

$$\beta_{ij}(H(G,\mathcal{P})) = \frac{\sum_{u\in V_i, v\in V_j}\alpha_u(G)\alpha_v(G)\beta_{uv}(G)}{\alpha(G[V_i])\alpha(G[V_j])}.$$

Note that $H(G,\mathcal{P})$ is invariant under scaling the nodeweights of G.

In the special case when G is unweighted, this definition specializes to

$$\alpha_i(H(G,\mathcal{P})) = \frac{|V_i|}{|V(G)|},$$

and

$$\beta_{ij}(H(G,\mathcal{P})) = \frac{e_G(V_i,V_j)}{|V_i|\cdot|V_j|}.$$

Here $e_G(V_i,V_j)$ denotes the number of edges $uv \in E(G)$ with $u \in V_i$ and $v \in V_j$; note that we allow $i = j$, in this case $e(V_i,V_i)$ is *twice* the number of edges spanned by V_i.

We denote by $\widehat{\mathcal{S}}_q(G)$ the set of factors of G with q nodes. Note that by our definition the factors are labeled graphs, but since permuting the nodes of a factor also gives a factor, we would not loose information by forgetting the labeling. We can consider $\widehat{\mathcal{S}}_q(G)$ as a subset of $\mathbb{R}^{q\times(q+1)}$.

We extend these definitions to functions $U \in \mathcal{W}_0$. Let $\mathcal{P}$ be a q-partition of $[0,1]$; we then define a weighted graph $H(U,\mathcal{P})$ on $V(H(U,\mathcal{P})) = [q]$, where node i has weight

$$\alpha_i(H(U,\mathcal{P})) = \lambda(V_i),$$

and edge ij has weight

$$\beta_{ij}(H(U,\mathcal{P})) = \frac{1}{\lambda(V_i)\lambda(V_j)} \int_{V_i\times V_j} U(x,y)\,dx\,dy.$$

(If $\lambda(V_i)\lambda(V_j) = 0$, then we define $\beta_{ij}(H(U,\mathcal{P})) = 0$.) We call the graph $H(U,\mathcal{P})$ a *factor* of U, and use the symbol $\widehat{\mathcal{S}}_q(U)$ to denote the set of all factors of U with q nodes.

Note that the knowledge of the factors of G is enough to recover the ground state energies of G. Indeed, in terms of the factors of G, the ground state energy defined in (38) can be expressed as

$$\mathcal{E}(G,B) = \max_{(a,X)\in\widehat{\mathcal{S}}_q(G)} \sum_{i,j=1}^{q} a_i a_j B_{ij} X_{ij}, \tag{43}$$

where $a = (a_1,\dots,a_q)$ and $X = (X_{ij})_{1\le i,j\le q}$.

6.4 Fractional Factor Graphs

The set $\widehat{\mathcal{S}}_q(G)$ is typically a very large finite set, which makes it difficult to work with. It will be convenient to introduce a fractional version of factors.

Let G be a weighted graph. For every fractional partition $\rho = (\rho_1,\dots,\rho_q)$ of $V(G)$, we define the *fractional factor* G_ρ as the graph with nodeweights

$$\alpha_i(G_\rho) = \frac{\sum_{u\in V(G)} \rho_i(u)\alpha_u(G)}{\alpha(G)},$$

and edgeweights

$$\beta_{ij}(G_\rho) = \frac{\sum_{u,v\in V} \rho_i(u)\rho_j(v)\alpha_u(G)\alpha_v(G)\beta_{uv}(G)}{\alpha_i(G_\rho)\alpha_j(G_\rho)}.$$

Let $\mathcal{S}_q(G)$ denote the set of all fractional factors of G with q nodes. Then $\mathcal{S}_q(G)$ is a closed set; it is not convex in general.

We also extend these notions to functions. Let $U \in \mathcal{W}_0$ and let $\rho = (\rho_1,\dots,\rho_q)$ be a fractional partition of $[0,1]$. Set $\alpha_i(\rho) = \int_0^1 \rho_i(x)\,dx$. Then we define

$$\alpha_i(U_\rho) = \alpha_i(\rho),$$

and

$$\beta_{ij}(U_\rho) = \frac{1}{\alpha_i(\rho)\alpha_j(\rho)} \int_{[0,1]^2} \rho_i(x)\rho_j(y)U(x,y)\,dx\,dy.$$

Let $\mathcal{S}_q(U)$ denote the set of all fractional factors of U with q nodes. For a weighted graph G and a vector $a \in \mathsf{Pd}_q$, let $\widehat{\mathcal{B}}_a(G)$ denote the set of all weighted adjacency matrices of all factors of G with nodeweights $a_1, \dots, a_q$. So $\mathcal{S}_q(G)$ is the set of all pairs (a, B) with $B \in \mathcal{B}_a(G)$. The sets $\mathcal{B}_a(G)$, $\widehat{\mathcal{B}}_a(U)$ and $\mathcal{B}_a(U)$ are defined analogously.

We clearly have that $0 \le \beta_{ij}(U_\rho) \le 1$ whenever $U \in \mathcal{W}_0$. Furthermore, it is not hard to see that for $U \in \mathcal{W}_0$, the set $\mathcal{S}_q(U)$ is closed in the obvious topology of weighted labeled graphs on q nodes.

Lemma 6.7 ([BCLSV06⁺]). *For every $U \in \mathcal{W}_0$, the set $\mathcal{S}_q(U)$ is the closure of $\widehat{\mathcal{S}}_q(U)$.*

(The two sets are not equal in general.)

For every weighted graph G, $\mathcal{S}_q(G)$ is again a closed connected set. Obviously, $\mathcal{S}_q(G)$ contains $\widehat{\mathcal{S}}_q(G)$, but it is not its closure in general (since the latter is a finite set). It is not hard to see that for every weighted graph G,

$$\mathcal{S}_q(G) = \mathcal{S}_q(W_G) = \widehat{\mathcal{S}}_q(W_G). \tag{44}$$

Clearly $\widehat{\mathcal{S}}_q(G)$ is a finite subset of these infinite sets. But it can be shown that it is not much smaller:

Lemma 6.8 ([BCLSV06⁺]). *If c is the largest nodeweight in G, then for every $H \in \mathcal{S}_q(G)$ there is an $H' \in \widehat{\mathcal{S}}_q(G)$ such that $\delta_\square(H, H') \le 4q^2\sqrt{c}$.*

Most of the time, we will work with the fractional versions, which are much easier to handle.

6.5 Microcanonical Ground State Energy, a.k.a. Multiway Cut Problems

We have seen that convergence of the sequence (G_n) implies convergence of the free energies and ground state energies. But the converse does not hold. To get convergence of (G_n) from the convergence of the ground state energies, we will need a finer measure than ground state energy, where the sizes of the partition classes are also taken into account (in physics terms, this could be called the "microcanonical version" of the ground state energy). We will see that this quantity also contains a number of frequently studied graph parameters.

For a finite set S, $q \ge 1$ and $a \in \mathsf{Pd}_q$, let a^S denote the set of all maps $\phi :\ S \to [q]$ such that

$$|\phi^{-1}(i) - a_i|S|| \le 1 \tag{45}$$

for every $i \in [q]$. (In other words, we prescribe the proportions of elements of S mapped onto each $i \in [q]$, as closely as possible.) For a simple graph G and a softcore graph H with node weights one, we then introduce microcanonical homomorphism numbers

$$\hom_a(G,H) = \sum_{\phi \in a^{V(G)}} \hom_\phi(G,H).$$

The *microcanonical ground state energies* and *free energies* of the model B on a graph G with n nodes are then defined as

$$\widehat{\mathcal{E}}_a(G,B) = \frac{1}{n^2} \max_{\phi \in a^{V(G)}} \sum_{ij \in E(G)} B_{\phi(i),\phi(j)}, \tag{46}$$

and

$$\widehat{\mathcal{P}}_a(G,B) = \frac{1}{n} \log \hom_a\big(\frac{1}{n} G, H\big) \tag{47}$$

respectively.

The ground state energy $\hat{\mathcal{E}}_a$ contains a number of important graph parameters as special cases. If

$$q = 2, \quad a = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1\,0 \\ 0\,0 \end{pmatrix} \tag{48}$$

then $\mathcal{E}_a(G,B)$ corresponds to the *densest subgraph* on $\alpha|V(G)|$ nodes. If

$$q = 2, \quad a = \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix} \tag{49}$$

then $-\mathcal{E}_a(G,B)$ is the *minimal bisection*; if we replace here B by $-B$, then $\mathcal{E}_a(G,B)$ is the *maximal bisection*. For $q > 2$, we get *multiway cut problems* in a similar way.

The free energy $\hat{\mathcal{P}}_a$ is a finer measure. For example, in the case (48), $\mathcal{P}_a(G,B)$ will pick out the density of an induced subgraph that is not necessarily the maximum, but for which the number of induced subgraphs with this density is large, at the cost of some loss in density.

The use of (45) is cumbersome and in some cases it leads to unpleasant discontinuities; it will be much more convenient to work with a fractional version. We formulate our definition for a weighted graph G. Then for all $a \in \mathsf{Pd}_q$ and $B \in \mathsf{Sym}_q$ we define

$$\mathcal{E}_a(G,B) = \max_\rho \frac{1}{2} \sum_{u,v \in V(G)} \alpha_u(G)\alpha_v(G)\beta_{uv}(G) \sum_{i,j \in [q]} \rho_i(u)\rho_j(v)B_{ij}, \tag{50}$$

where ρ ranges over all fractional q-partitions of $V(G)$ such that

$$\sum_{u \in V(G)} \alpha_v(G)\rho_i(v) = a_i.$$

We also extend the notion of microcanonical ground state energy to functions. For every $W \in \mathcal{W}_0$, $a \in \mathsf{Pd}_q$, and $B \in \mathsf{Sym}_q$, we define

$$\mathcal{E}_a(W,B) = \max_{\rho:\alpha_\rho=a} \frac{1}{2} \sum_{i,j=1}^{q} B_{ij} \int_{[0,1]^2} \rho_i(x)\rho_j(y)W(x,y)\,dx\,dy, \tag{51}$$

where ρ ranges over fractional q-partitions of $[0,1]$ with $\alpha_i(\rho) = a_i$ for all i. It is easy to see that

$$\mathcal{E}_a(G,B) = \mathcal{E}_a(W_G,B).$$

Using a result from [AVKK03], it follows easily that for every simple graph G on n nodes and every matrix B,

$$|\widehat{\mathcal{E}}_a(G,B) - \mathcal{E}_a(G,B)| \le \frac{6q^3}{n}\|B\|_\infty. \tag{52}$$

The notion of the microcanonical free energy for functions is defined analogously: for every $W \in \mathcal{W}_0$, $a \in \mathsf{Pd}_q$, and $B \in \mathsf{Sym}_q$, we define

$$\mathcal{P}_a(W,B) = \max_{\rho:\alpha(\rho)=a} \Big(\mathcal{H}(\rho) + \frac{1}{2}\sum_{i,j} B_{ij} \int_{[0,1]^2} W(x,y)\rho_i(x)\rho_j(y)\,dx\,dy\Big), \tag{53}$$

where ρ ranges over fractional q-partitions of $[0,1]$ with $\alpha_i(\rho) = a_i$ for all i.

6.6 Convergence from the Right

Our goal is to give a characterization of convergent graph sequences in terms of partition information, specifically the sets $\mathcal{S}_q(G)$ and the functions $\mathcal{E}_a(G,B)$. Recall that for two sets A, B in a metric space (X,d), the *Hausdorff distance* is defined as

$$d^{\mathrm{Hf}}(A,B) = \max(\sup_{x\in A}\inf_{y\in B} d(x,y), \sup_{x\in B}\inf_{y\in A} d(x,y)).$$

It turns out that the following three types of information about two graphs G, G' are equivalent: (1) G and G' are close in the $\delta_\square$ distance; (2) $\mathcal{S}_q(G)$ and $\mathcal{S}_q(G')$ are close in the d^{Hf} distance for every q up to a certain bound, and (3) $|\mathcal{E}_a(G,B) - \mathcal{E}_a(G',B)|$ is small for all $a \in \mathsf{Pd}_q$ and $B \in \mathsf{Sym}_q$ for every q up to a certain bound.

The exact statement is the following:

Theorem 6.9 ([BCLSV06$^+$]).

(a) *For two simple graphs G, G' and $a \in \mathsf{Pd}_q$,*

$$d^{\mathrm{Hf}}_\square(\mathcal{S}_q(G), \mathcal{S}_q(G')) \le \delta_\square(G,G')$$

(b) *For two simple graphs G, G', $a \in \mathsf{Pd}_q$ and $B \in \mathsf{Sym}_q$,*

$$|\mathcal{E}_a(G,B) - \mathcal{E}_a(G',B)| \le q^2 \delta_\square^{\mathrm{Hf}}(\mathcal{S}_q(G), \mathcal{S}_q(G')).$$

(c) *Let G, G' be two simple graphs, and suppose that*

$$|\mathcal{E}_a(G,B) - \mathcal{E}_a(G',B)| \le c_3 \frac{\varepsilon^2}{q^2}$$

for all $q \le 4 \cdot 2^{c_4/\varepsilon^2}$, $a \in \mathsf{Pd}_q$ and $B \in \mathsf{Sym}_q$. Then $\delta_\square(G,G') < \varepsilon$.

Similar statements hold for the microcanonical free energy, see [BCLSV06$^+$].

6.7 Summary: Convergence Criteria

The following theorem summarizes several equivalent conditions for a sequence of (dense) graphs to be convergent:

Theorem 6.10 ([BCLSV06$^+$]). *Let (G_n) be a sequence of simple graphs with $|V(G_n)| \to \infty$. Then the following are equivalent:*

(a) *For every simple graph F, $t(F, G_n)$ is convergent.*
(b) *The sequence (G_n) is Cauchy in the $\delta_\square$ metric.*
(c) *For every $q \ge 1$, the sequence $\mathcal{S}_q(G_n)$ is Cauchy with respect to the Hausdorff metric d_1^{Hf}.*
(d) *For every $q \ge 1$, $a \in \mathsf{Pd}_q$ and $B \in \mathsf{Sym}_q$, the sequence $\mathcal{E}_a(G_n, B)$ is a Cauchy sequence.*
(e) *For every $q \ge 1$, $a \in \mathsf{Pd}_q$ and $B \in \mathsf{Sym}_q$, the sequence $\widehat{\mathcal{P}}_a(G_n, B)$ is a Cauchy sequence.*

We can also characterize convergent graph sequences in terms of Szemerédi partitions.

Supplement 6.11 ([BCLSV06$^+$]). *The following two conditions are also equivalent to conditions* (a)–(e) *in Theorem* 6.10*:*

(f) *For every $k \ge 1$ there is an $n_k \ge 1$ such that if $n, m > n_k$, then G_n and G_m have weak Szemerédi k-partitions $\mathcal{P}$ and $\mathcal{P}'$ with error $2/\sqrt{\log k}$ such that $d_\square(H(G_n,\mathcal{P}), H(G_m,\mathcal{P}')) < 2/\sqrt{\log k}$.*
(g) *For every $k \ge 1$ there is an $n_k \ge 1$ such that if $n, m > n_k$, then G_n and G_m have strong Szemerédi k-partitions $\mathcal{P}$ and $\mathcal{P}'$ with error $1/\log^* k$ such that $d_\square(H(G_n,\mathcal{P}), H(G_m,\mathcal{P}')) < 1/\log^* k$.*

A convergent sequence has a limit $W \in \mathcal{W}_0$ by Theorem 4.6. The conditions in Theorem 6.10 can be rephrased to characterize the convergence to this limit:

Theorem 6.12 ([BCLSV06$^+$]). *For a sequence (G_n) of simple graphs with $|V(G_n)| \to \infty$, and for any $W \in \mathcal{W}_0$, the following are equivalent:*

(a) *For every simple graph* F, $t(F,G_n) \to t(F,W)$.
(b) $\delta_\square(G_n,W) \to 0$.
(c) *For every* $q \geq 1$, $\mathcal{S}_q(G_n) \to \mathcal{S}_q(W)$ *in the Hausdorff metric* d_1^{Hf}.
(d) *For every* $q \geq 1$, $a \in \mathsf{Pd}_q$ *and* $B \in \mathsf{Sym}_q$, $\mathcal{E}_a(G_n,B) \to \mathcal{E}_a(W,B)$.
(e) *For every* $q \geq 1$, $a \in \mathsf{Pd}_q$ *and* $B \in \mathsf{Sym}_q$, $\widehat{\mathcal{P}}_a(G_n,H) \to \mathcal{P}_a(W,B)$.

One can define Szemerédi partitions for the limit objects $W \in \mathcal{W}_0$ (see [LSz05a]), and then formulate analogues of (f) and (g) in Supplement 6.11 describing the convergence to the limit.

7 Homomorphisms and Extremal Graph Theory

7.1 Inequalities between Homomorphism Numbers

We have mentioned in the introduction that many results in extremal graph theory can be expressed as algebraic inequalities between homomorphism densities. Every algebraic inequality that holds for all finite graphs also holds for simple graph parameters in the closure $\mathcal{T}_0$ (and of course vice versa). So for example, Goodman's Theorem (1) is equivalent to saying that for every simple graph parameter $t \in \mathcal{T}_0$,

$$t(K_3) \geq t(K_2)(2t(K_2) - 1). \tag{54}$$

By multiplicativity, every algebraic inequality is equivalent to a linear inequality. For example, (1) is equivalent to

$$t(K_3) \geq 2t(K_2K_2) - t(K_2).$$

The positive semidefiniteness of the connection matrix (reflection positivity) implies linear and nonlinear inequalities between the values of a simple graph parameters $t \in \mathcal{T}_0$ for "small" graphs. In fact, Theorem 4.5 says that every (say, algebraic) inequality between the values of simple graph parameters $t \in \mathcal{T}_0$ is a consequence of multiplicativity and reflection positivity.

Let us describe some of these derivations. Fix a simple graph parameter $t \in \mathcal{T}_0$. If $v = (v_G)$ is any real vector with finite support, indexed by k-labeled graphs, then

$$v^{\mathsf{T}} M(t,k) v \geq 0 \tag{55}$$

gives a linear inequality between the values $t(F)$.

Probably all linear inequalities can be obtained by taking nonnegative linear combinations of inequalities (55). However, the (infinitely many) inequalities (55) and their consequences are not so easy to understand, and we formulate some special inequalities.

Trivial linear inequalities are that if F_1 is a subgraph of F_2 (not necessarily on the same set of nodes), then

$$t(F_1) \geq t(F_2). \tag{56}$$

For any two graphs $F_1 \subseteq F_2$ on the same set of nodes, the expression

$$\sum_{F_1 \subseteq F \subseteq F_2} (-1)^{|E(F_2)\setminus E(F)|} \hom(F, G)$$

counts (by inclusion-exclusion) the number of homomorphisms of F_1 into G that map the edges in $E(F_2) \setminus E(F_1)$ onto non-adjacent pairs. This number is nonnegative, which implies that

$$\sum_{F_1 \subseteq F \subseteq F_2} (-1)^{|E(F_2)\setminus E(F)|} t(F, G) \geq 0. \tag{57}$$

If we fix $V(F) = V$ with $|V| = k$, then $t(F)$ can be considered as a setfunction on $\binom{k}{2}$ elements. Then (57) can be used to show that this setfunction is supermodular, i.e., it satisfies

$$t(F_1 \cup F_2) + t(F_1 \cap F_2) \geq t(F_1) \cap t(F_2). \tag{58}$$

We leave it to the reader as an exercise to derive these inequalities from (55).

Semidefiniteness of the connection matrix can also be formulated in terms of nonlinear inequalities (nonnegativity of certain determinants). One of these is worth mentioning. Let G be a k-labeled graph, then for $t \in \mathcal{T}_0$, we have

$$t(GG) \geq t(G)^2.$$

As special cases, we mention that for the path P_3 on 3 nodes,

$$t(P_3) \geq t(K_2)^2 \tag{59}$$

and

$$t(C_4) \geq t(P_3)^2. \tag{60}$$

7.2 Re-proving Some Results in Extremal Graph Theory

Let us start with deriving (54). By (57),

$$t(K_3) - 2t(P_3) + t(K_2K_1) \geq 0,$$

Using multiplicativity, we have $t(K_2 + K_1) = t(K_2)t(K_1) = t(K_2)$. Furthermore, by (59) we have $t(P_3) \geq t(K_2)^2$. This implies (54).

We can derive the following theorem of Moon and Moser [MM62] in a similar way: Let G be a graph with n nodes, and let N_r denote the number of complete subgraphs with r nodes. Then

$$\frac{N_{r+1}}{N_r} \geq \frac{1}{r^2 - 1}\left(r^2 \frac{N_r}{N_{r-1}} - n\right). \tag{61}$$

Using that $N_r = t(K_r, G)n^r/r!$, this inequality can be expressed in the following simpler form: For every $t \in \mathcal{T}_0$,

$$r\frac{t(K_r)}{t(K_{r-1})} \le (r-1)\frac{t(K_{r+1})}{t(K_r)} + 1. \qquad (62)$$

This shows that (54) is a special case, and the derivation of (54) above can be extended.

The (simplest, asymptotic) case of the Kruskal-Katona theorem,

$$t(K_3)^2 \le t(K_2)^3, \qquad (63)$$

also follows. One uses multiplicativity to write it as $t(K_3)^2 \le t(K_2)t(K_2K_2)$; by monotonicity, it suffices to prove the stronger inequality $t(K_3)^2 \le t(K_2)t(C_4)$; which then follows by considering the following submatrix of M_2:

$$\begin{pmatrix} t(K_2) & t(K_3) \\ t(K_3) & t(C_4) \end{pmatrix}.$$

7.3 A 2-dimensional Projection

Perhaps the most basic question about inequalities between homomorphism densities concerns edges and triangles. What are the possible pairs $(t(K_2, G), t(K_3, G))$? If we disregard number theoretic conditions like these numbers must be rational, we can ask: What are the possible pairs $(t(K_2), t(K_3))$, where $t \in \mathcal{T}_0$? In other words, what is the projection $\mathcal{T}_0'$ of $\mathcal{T}_0$ onto the 2-dimensional plane determined by the density of edges and density of triangles?

The answer to this questions turns out quite complicated and is only partially solved. Figure 2 describes the set $\mathcal{T}_0'$. The upper boundary curve is given the edge-density is given by the Kruskal-Katona Theorem (63):

$$x^3 = y^2$$

(and this is tight for all densities). The lower bound is more complicated. Inequality (54) gives a parabola that is a lower bound. However, this is only tight for special values of the edge-density: $t(K_2) = 1 - 1/k$, $k = 1, 2, \dots$. For these values, a Turán graph (complete k-partite graph with equal color classes) gives equality. Bollobás [Bol75] proved that in the intervals between these values, the lower boundary is above the chord. A conjecture for the curve was formulated in [LSi83]: An extremal graph in the interval $1 - 1/k \le t(K_2) \le 1 - 1/(k+1)$ is a complete $(k+1)$-partite graph with $k-1$ equal color classes. The size of the two special color classes must be optimized to make sure that the density of triangles is minimized, subject to the given edge density. For the interval $1/2 \le t(K_2) \le 2/3$, the optimization yields the cubic curve

$$9y^2 - 18xy + 8x - 3x^2 + 6x^3 = 0.$$

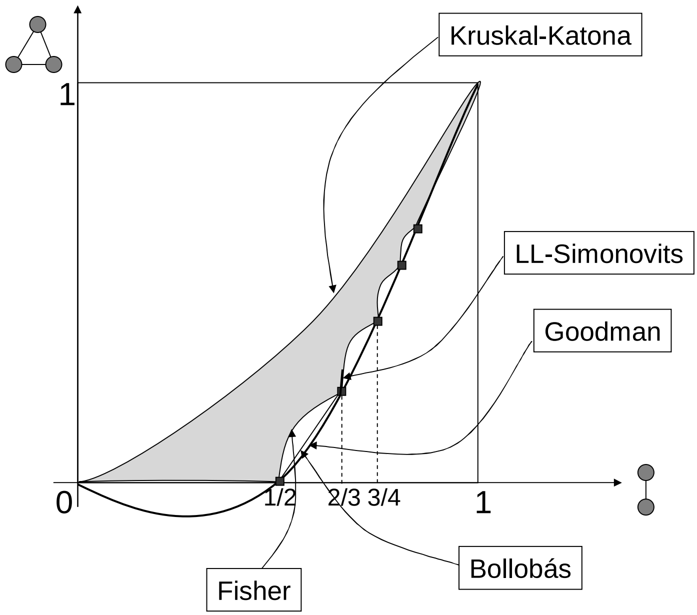


Fig. 2. The region of possible edge densities and triangle densities. (The concave arcs are distorted to make the qualitative properties more visible.)

This case was proved by Fisher [Fis89]; a recent proof by Razborov [Raz05] uses methods quite closely related to those described in this paper. The conjecture was proved in [LSi83] if $t(K_2)$ was in a small neighborhood of any one of the special values $1 - 1/k$. The general case is open.

One should add that this conjecture is all that is needed to describe $\mathcal{T}_0'$: it is easy to see that between two points of $\mathcal{T}_0'$ the plane on a vertical line, the whole interval is contained in $\mathcal{T}_0'$.

8 Graphs with Bounded Degree

Fix a positive integer D, and consider (as the middle graph G in (3) only graphs in which all degrees are at most D. The normalization $t(F,G)$ we used before does not make sense any more, since (if G is large) the probability that a random mapping from a given F is a homomorphism is very small.

How to normalize $\hom(F,G)$ in this case? Let $n = |V(G)|$ and $k = |V(F)|$. If F is not connected, then $\hom(F,G)$ is just the product of the numbers $\hom(F',G)$, where F' is a connected component of F, so we may restrict our attention to the case when F is connected; then there is the following rather obvious upper bound:

$$\hom(F, G) \le n \cdot D^{k-1}.$$

So it makes sense to consider

$$\tau(F, G) = \frac{\hom(F, G)}{n} \tag{64}$$

(Since we consider D and F fixed, it does not seem to help to divide by D^{k-1}.)

8.1 Convergence for Graphs with Bounded Degree

Let $(G_1, G_2, \dots)$ be a sequence of graphs whose degrees are uniformly bounded by D. We say that this sequence is *locally convergent*, or just *convergent*, if $\tau(F, G_n)$ tends to a limit for every connected graph F.

There is another way to define this. Given a graph G with all degrees bounded by D, and a positive integer r, let $S_G(v, r)$ denote the neighborhood of node v with radius r. We consider $S_G(v, r)$ as a rooted graph (where v is its root). For fixed D and r, there is a finite number of possible neighborhoods, and we can make a statistic of these: we denote by $\nu_G(N, r)$ the fraction of nodes $v \in V(G)$ for which $S_G(v, r) \cong N$. Then ν_G is a probability distribution on all possible r-neighborhoods.

For bounded degrees, it is easy to see that the probability distributions $\nu_{G_n}(\cdot, r)$ tend to a limit distribution for all $r > 0$ if and only if G_n converges in the sense defined above. The convergence of the probability distributions $\nu_{G_n}(\cdot, r)$ can therefore be used as an alternative characterization of convergent sequences $(G_1, G_2, \dots)$ of graphs with degrees bounded by D.

This notion of convergence was introduced implicitly by Aldous [Ald98], and explicitly by Benjamini and Schramm [BS01]. It was extended to the case of bounded average degree by Lyons [Lyo06$^+$]. The limit object has several descriptions, the strongest is due to Elek [Elek06$^+$]: A *continuous graphing* is an infinite graph on $[0, 1]$ that has the following structure: we take a finite number of continuous measure preserving involutions $\phi_1, \dots, \phi_N : [0, 1] \to [0, 1]$, and connect every $x \in [0, 1]$ to every $\phi_i(x)$ $(i = 1, \dots, N)$ such that $x \ne \phi_i(x)$ by an edge.

Every graphing defines a probability distribution on countable graphs with bounded degree with a specified root: we pick a uniform random point x in $[0, 1]$, and consider the connected component of the graphing containing x, with root x. This yields the description of the limit by Benjamini and Schramm.

8.2 Left and Right Convergence

The theory of convergence of graphs with bounded degree is much less satisfactory than the analogous theory of dense graphs described above. In particular, the "right" notion of distance and a powerful analogue of Szemerédi's Lemma are missing. But there are some nontrivial facts, relating homomorphisms into a large graph to homomorphisms from it, which can be thought of as analogues of some results in Section 6.

Let us start with an example.

Example 8.1. Consider the sequence of cycles C_n. It is trivial that this is convergent, but the numbers $\hom(C_n, K_2)$ alternate between 0 and 2.

But in a sense parity is all that goes wrong. Let us consider the subsequence C_{2n} of even cycles. Then for any graph H,

$$\hom(C_{2n}, H) = \sum_{i=1}^{k} \lambda_i^{2n},$$

where $\lambda_1 \geq \ldots \geq \lambda_k$ are the eigenvalues of H. Hence

$$\hom(C_{2n}, H)^{1/(2n)} \to \lambda_1.$$

It is perhaps more usual to take the logarithm here, to get the sequence $\frac{1}{2n} \log \hom(C_{2n}, H)$ (in statistical physics, this parameter is called the free energy or pressure of the H-colorings of C_{2n}). This sequence is convergent for every H, and the limiting parameter is $\log \lambda_1(H)$.

The sequence of odd cycles behaves similarly, but the limit parameter $a(H) = \lim \frac{1}{2n+1} \log \hom(C_{2n+1}, H)$ is a bit more complicated to describe: $a(H)$ is the logarithm of largest eigenvalue of its non-bipartite components ($a(H) = -\infty$ if H is bipartite).

Based on this example, let us consider a convergent sequence (G_n) of graphs with bounded degree, and ask for which graphs H the numbers

$$\widehat{\mathcal{P}}(G_n, H) = \frac{1}{|V(G_n)|} \log \hom(G_n, H)$$

converge.

Before discussing this further, let us rewrite the right hand side using Möbius inversion. Let

$$\psi(G, H) = \sum_{V \subset V(G)} (-1)^{|V(G)|-|V|} \log \hom(G[V], H).$$

Then

$$\log \hom(G, H) = \sum_{V \subset V(G)} \psi(G[V], H) = \sum_F \frac{\operatorname{ind}(F, G)}{|V(F)|!} \psi(F, H), \tag{65}$$

where the sum goes over all finite simple graphs F and $\operatorname{ind}(F, G)$ is the number of embedding of F into G as an induced subgraph. We thus have shown that

$$\widehat{\mathcal{P}}(G, H) = \sum_F \frac{\tau_{\mathrm{ind}}(F, G)}{|V(F)|!} \psi(F, H) \tag{66}$$

where

$$\tau_{\mathrm{ind}}(F,G) = \frac{1}{|V(G)|}\mathrm{ind}(F,G). \tag{67}$$

Recall that the numbers $\mathrm{ind}(F,G)$ can be obtained from the homomorphism numbers $\hom(F,G)$ by Möbius inversion (see (14) and (15)). If G_n is convergent in the sense that the normalized homomorphism numbers $\tau(F,G_n)$ are convergent for all F, then the numbers $\tau_{\mathrm{ind}}(F,G_n)$ are convergent as well; let $\tau_{\mathrm{ind}}(F,G_n) \to \tau(F)$. One might therefore hope that convergence of the sequence G_n implies convergence of the free energies $\widehat{\mathcal{P}}(G_n,H)$, with the limit given by the infinite sum

$$\mathcal{P}(\tau,H) = \sum_F \frac{\tau(F)}{|V(F)|!}\psi(F,H). \tag{68}$$

It is clear, however, that this cannot be true in general; an easy counterexample is the example $G_n = C_n$ discussed above. But it turns out that we can prove the convergence under suitable conditions on H and the maximal degree D of the graphs G_n.

Theorem 8.2 ([BCKLS06+]). *If (G_n) is a sequence of graphs with all degrees bounded by a constant D, and (G_n) is convergent, then $\widehat{\mathcal{P}}(G_n,H)$ has a limit for every unweighted target graph H in which all degrees are at least $\left(1-\frac{1}{2D}\right)|V(H)|$.*

The conceptually most transparent proof of the theorem (with a slightly worse constant) proceeds by proving uniform convergence of the expansion (68) using the method of cluster expansions. As stated, the theorem can be proven using Dobrushin's uniqueness theorem for uniform H-colorings on G_n.

The cluster expansion proof also gives an analogue of Theorem 8.2 for weighted graphs (see [BCKLS06+]) and allows to prove that convergence from the right for weighted graphs implies convergence from the left. In fact, we only need convergence from the right for graphs H that are small perturbations of completely looped complete graphs with edge weights 1 to conclude that G_n is convergent from the left.

Theorem 8.3 ([BCKLS06+]). *If (G_n) is a sequence of graphs with all degrees bounded by a constant D, and $\mathcal{P}(G_n,H)$ is convergent for every weighted graph H which is a looped complete graph with all edgeweights arbitrarily close to 1, then (G_n) is convergent.*

References

[Ald98] D. J. Aldous: Tree-valued Markov chains and Poisson-Galton-Watson distributions, in: *Microsurveys in Discrete Probability* (D. Aldous and J. Propp, editors), DIMACS Ser. Discrete Math. Theoret. Comput. Sci. **41** (1998) Amer. Math. Soc., Providence, RI. (1998), 1–20.

[AVKK03] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski: Random sampling and approximation of MAX-CSPs, *J. Comput. System Sci.* **67** (2003), 212–243.

[AN06+] N. Alon and A. Naor: Approximating the Cut-Norm via Grothendieck's Inequality, preprint: `http://research.microsoft.com/research/theory/naor/homepage%20files/cutnorm.pdf`

[AS06+] N. Alon and A. Shapira, Every monotone graph property is testable, Proc. of the 37th ACM STOC, Baltimore, ACM Press (2005), to appear: `http://www.math.tau.ac.il/~nogaa/PDFS/MonotoneSTOC.pdf`

[BK06+a] E. Babson and D. Kozlov: Complexes of graph homomorphisms, arXive: `http://lanl.arxiv.org/abs/math.CO/0310056`

[BK06+b] E. Babson and D. Kozlov: Proof of the Lovász Conjecture, arXive: `http://lanl.arxiv.org/abs/math.CO/0402395`

[BA99] A. Barabasi and R. Albert: Emergence of scaling in random networks, *Science* **286** (1999) 509–512.

[BR03] B. Bollobás and O. Riordan: Mathematical results on scale-free random graphs, in: *Handbook of graphs and networks*, 1–34, Wiley-VCH, Weinheim, 2003.

[BBCKK04] M. Biskup, C. Borgs, J. T. Chayes, L. Kleinwaks and R. Kotecký: Partition function zeros at first-order phase transitions: A general analysis, *Commun. Math. Phys.* **251** (2004) 79–131.

[BRST01] B. Bollobás, O. M. Riordan, J. Spencer, and G. Tusnády: The degree sequence of a scale-free random graph process, *Random Structures and Algorithms* **18** (2001), 279–290.

[Bol75] B. Bollobás: Relations between sets of complete subgraphs, in: *Combinatorics*, Proc. 5th British Comb. Conf. (ed. C.St.J.A. Nash-Williams, J. Sheehan), Utilitas Math. (1975), 79–84.

[BBCR03] B. Bollobás, C. Borgs, J. T. Chayes, and O. Riordan: Directed scale-free graphs, in: *Proceedings of the 14th ACM-SIAM Symposium on Discrete Algorithms*, 132–139, 2003.

[Bor06+] C. Borgs: Statistical Physics Expansion Methods for Combinatorics and Computer Science, CBMS lecture notes (in preparation).

[BCL06+] C. Borgs and J. Chayes and L. Lovász: Unique limits of dense graph sequences (in preparation).

[BCKLS06+] C. Borgs, J. Chayes, J. Kahn, L. Lovász and V. T. Sós: Convergent sequences for sparse graphs (in preparation).

[BCLSV06+] C. Borgs, J. Chayes, L. Lovász, V. T. Sós and K. Vesztergombi: Convergent sequences of dense graphs (in preparation).

[BS01] I. Benjamini and O. Schramm: Recurrence of Distributional Limits of Finite Planar Graphs, *Electronic J. Probab.* **6** (2001), no. 23, 1–13.

[Choi83] M.-D. Choi: Tricks or Treats with the Hilbert Matrix, *Amer. Math. Monthly* **90** (1983), 301–312.

[CGW89] F. Chung, R.L. Graham and R.M. Wilson: Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[CF03] C. Cooper and A. Frieze: On a general model of web graphs, *Rand. Struct. Alg.* **22** (2003) 311–335.

[Elek06+] G. Elek: Graphings and graph sequences (preprint).

[KW82] K. Drühl and H. Wagner: Algebraic formulation of duality transforms for abelian lattice models, *Anals of Physics* **141** (1982), 225–253.

[ELS79] P. Erdös and L. Lovász, J. Spencer: Strong independence of graph-copy functions, in: *Graph Theory and Related Topics*, Academic Press (1979), 165–172.

[Fis89] D. C. Fisher: Lower bounds on the number of triangles in a graph, *J. Graph Theory* **13** (1989), 505–512.

[FR90] D. C. Fisher and J. Ryan: Conjectures on the number of complete subgraphs, in: *Proc. of the 20-th Southeastern Conf. on Comb., Graph Theory, and Computing*, Congr. Numer. **70** (1990), 217–219.

[FS90] D. C. Fisher and A. Solow: Dependence polynomials, *Discrete Math.* **82** (1990), 251–258.

[FLW] M. Freedman and L. Lovász, D. Welsh (unpublished).

[FLS04] M. Freedman, L. Lovász and A. Schrijver: Reflection positivity, rank connectivity, and homomorphism of graphs (MSR Tech Report # MSR-TR-2004-41).
`ftp://ftp.research.microsoft.com/pub/tr/TR-2004-41.pdf`

[FK99] A. Frieze and R. Kannan: Quick approximation to matrices and applications, *Combinatorica* **19** (1999), 175–220.

[Goo59] A. W. Goodman: On sets of aquaintences and strangers at any party, *Amer. Math. Monthly* **66** (1959) 778–783.

[GGR98] O. Goldreich, S. Goldwasser and D. Ron: Property testing and its connection to learning and approximation, *J. ACM* **45** (1998), 653–750.

[Gow97] W. T. Gowers: Lower bounds of tower type for Szemerédi's Uniformity Lemma, *Geom. Func. Anal.* **7** (1997), 322–337.

[HJ93] P. de la Harpe and V. F. R. Jones: Graph Invariants Related to Statistical Mechanical Models: Examples and Problems, *Journal of Combinatorial Theory B* **57** (1993), 207–227.

[HN04] P. Hell and J. Nešetřil: *Graphs and Homomorphisms*, Oxford University Press, 2004.

[KS96] J. Komlós and M. Simonovits: Szemerédi's Regularity Lemma and its applications in graph theory, in: *Combinatorics, Paul Erdos is Eighty* (D. Miklos et. al, eds.), Bolyai Society Mathematical Studies **2** (1996), pp. 295–352.

[LY52] T. D. Lee and C. N. Yang, Statistical theory of equations of state and phase transitions: II. Lattice gas and Ising model, *Phys. Rev.* **87** (1952) 410–419.

[Lov67] L. Lovász: Operations with structures, *Acta Math. Hung.* **18** (1967), 321–328.

[Lov72] L. Lovász: Direct product in locally finite categories, *Acta Sci. Math. Szeged* **23** (1972), 319–322.

[Lov06+] L. Lovász: The rank of connection matrices and the dimension of graph algebras, *Europ. J. Combin.*, to appear:
`http://research.microsoft.com/~lovasz/homdim.pdf`

[LSc06+] L. Lovász and A. Schrijver: Graph parameters and semigroup functions (manuscript).

[LSi76] L. Lovász and M. Simonovits: On the number of complete subgraphs of a graph (M. Simonovits), in: *Combinatorics*, Proc. 5th British Comb. Conf. (ed. C.St.J.A. Nash-Williams, J.Sheehan), Utilitas Math. (1976), 439–441.

[LSi83] L. Lovász and M. Simonovits: On the number of complete subgraphs of a graph II, in: *Studies in Pure Math.*, To the memory of P. Turán (ed. P. Erdös), Akadémiai Kiadó (1983), 459–495.

[LSó06+] L. Lovász and V. T. Sós: Generalized quasirandom graphs, preprint: `http://research.microsoft.com/~lovasz/quasirandom4.pdf`

[LSz04] L. Lovász and B. Szegedy: Limits of dense graph sequences, Microsoft Research Technical Report MSR-TR-2004-79: `ftp://ftp.research.microsoft.com/pub/tr/TR-2004-79.pdf`

[LSz05a] L. Lovász and B. Szegedy: Szemerédi's Lemma for the analyst, Microsoft Research Technical Report MSR-TR-2005-90: `ftp://ftp.research.microsoft.com/pub/tr/TR-2005-90.pdf`

[LSz05b] L. Lovász and B. Szegedy: Contractors and connectors of graph algebras Microsoft Research Technical Report TR-2005-91: `ftp://ftp.research.microsoft.com/pub/tr/TR-2005-91.pdf`

[LSz05c] L. Lovász and B. Szegedy: Graph limits and testing hereditary graph properties, Microsoft Research Technical Report MSR-TR-2005-110: `ftp://ftp.research.microsoft.com/pub/tr/TR-2005-110.pdf`

[LSz06+] L. Lovász and B. Szegedy: Moments of 2-variable functions (in preparation).

[Lyo06+] R. Lyons: Asymptotic Enumeration of Spanning Trees, preprint: `http://mypage.iu.edu/~rdlyons/pdf/est.pdf`

[MSS93] H. Mahmound, R. Smythe and J. Szymański: On the structure of plane-oriented trees and their branches. *Random Struct. Alg.* **3** (1993) 255–266.

[Mat03] J. Matoušek: *Using the Borsuk-Ulam Theorem: Lectures on Topological Methods in Combinatorics and Geometry* Springer, 2003.

[MM62] J. W. Moon, L. Moser: On a problem of Turn, *Magyar Tud. Akad. Mat. Kutat Int. Kzl.* **7** (1962), 283–286.

[Pri76] D. J. de S. Price: A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, **27** (1976) 292–306.

[Raz05] A. Razborov (unpublished).

[SS91] M. Simonovits and V.T. Sós: Szemerédi's partition and quasirandomness, *Random Structures Algorithms* **2** (1991), 1–10.

[SS97] M. Simonovits and V. T. Sós: Hereditarily extended properties, quasirandom graphs and not necessarily induced subgraphs. *Combinatorica* **17** (1997), 577–596.

[SS03] M. Simonovits and V. T. Sós: Hereditary extended properties, quasirandom graphs and induced subgraphs, *Combinatorics, Probability and Computing 12* (2003), 319–344.

[Sze05] B. Szegedy: Edge models and reflection positivity, preprint: http://arxiv.org/abs/math.CO/0505035

[Tho87] A. Thomason: Pseudorandom graphs, in: *Random graphs '85* North-Holland Math. Stud. **144**, North-Holland, Amsterdam, 1987, 307–331.

[YL52] C. N. Yang and T. D. Lee: Statistical theory of equations of state and phase transitions: I. Theory of condensation, *Phys. Rev.* **87** (1952), 404–409.

[Yule24] G. U. Yule: A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis. *Philos. Trans. Roy. Soc. London*, Ser. B **213** (1924) 21–87.

[Whi32] H. Whitney: The coloring of graphs, *Ann. of Math.* **33** (1932), 688–718.

Efficient Algorithms for Parameterized H-colorings*

Josep Díaz, Maria Serna, and Dimitrios M. Thilikos**

ALBCOM Research Group, Dept. de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Campus Nord–Edifici Ω, E-08034 Barcelona, Spain
{diaz,mjserna,sedthilk}@lsi.upc.es

Summary. We study the fixed parameter tractability of the restrictive H-coloring and the restrictive list H-coloring problems, introduced in [DST01]. The parameterizations are defined by fixing the number of pre-images of a subset C of the vertices in H through a partial weight assignment (C, K). We define two families of partially weighted graphs: the *simple* and the *plain*. For the class of simple partially weighted graphs, we show the fixed parameter tractability of the list (H, C, K)-coloring problem. For the more general class of plain partially weighted graphs, we prove the fixed parameter tractability of the (H, C, K)-coloring problem.

AMS Subject Classification. 68R10, 05C15.
Keywords. Restrictive H-coloring, parametrization, FPT, NP-complete.

1 Introduction

Given two graphs G and H, a *homomorphism* from G to H is any function mapping the vertices in G to vertices in H, in such a way that the image of an edge is also an edge. In the case that H is fixed, such a homomorphism is called an H*-coloring* of G. The more general version in which a list of allowed colors (vertices of H) is given for each vertex is known as a *list* H*-coloring.* See [HN04] for further extensive background on morphisms on graphs and in particular on H-coloring.

* An extended abstract containing some of the results in this paper was presented at the *12th Annual European Symposium on Algorithms* (ESA 2004).
** The work of the first author was partially supported by the Distinció per a la recerca of the Generalitat de Catalunya 2002. The work of the first and third authors was partially supported by the Spanish CICYT project TIN-2004-0795 (GRAMMARS). The work of the second author was partially supported by the Spanish CICYT project TIN-2005-09918-C02-02 (ASCE).

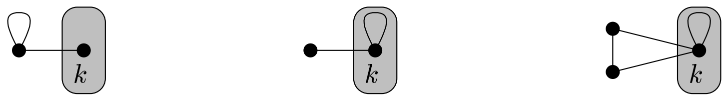
For a fixed graph H, the *H-coloring problem* asks whether there exists an H-coloring of the input graph G and the *list H-coloring problem* asks whether there exists a list H-coloring of the input graph G. The complexity of the H-coloring problem was studied in [HN90], where the following complexity dichotomy was proved: the H-coloring problem belongs to P if H is bipartite or contains a loop, otherwise the problem is NP-Complete. For the list H-coloring problem, there is also a dichotomy: the problem is in P if H is a bi-arc graph, otherwise the problem is NP-Complete [FHH03].

The case of an H-coloring in which, in addition, the number of pre-images of the vertices in H are restricted, is known as a *restrictive H-coloring* or a *restrictive list H-coloring*. The corresponding decision problems are known as the *restrictive H-coloring* and the *restrictive list H-coloring* problems. The complexity of those problems was studied in [DST05], where it was proved that the problems are in P if all the connected components of H are either complete reflexive graphs or complete irreflexive bipartite graphs, otherwise the problems are NP-Complete.

To study the structural characteristics that make a problem difficult to solve, a fruitful approach has been to divide the input of a problem into two parts one of them, the *non-parameterized*, remains as part of the input while the other one, the *parameterized*, is fixed independently of the input. This line of research is known as *parameterized complexity*, see for ex. [DF99]. More specifically, the input to a parameterized problem, is a pair (S, K) where S is the main part and K is the parameterized part. Typically, K is an integer, a sequence of integers, or some other structure, that is fixed independently of the input, and thus it is called the *parameter*. The parameterized version of the *restrictive H-coloring* problems in which the number of allowed pre-images of some vertices in H are fixed independently of the input is known as the *(H, C, K)-coloring* and was introduced in [DST01]. In the triple (H, C, K), C is the set of vertices in H with restrictions and K is a weight assignment on C defining the number of pre-images of these vertices in G. We will say that (H, C, K) is the *partial weight assignment* corresponding to the parameterized part of the restrictive H-coloring problem.

The considered parameterization allows us to capture some well known parameterized problems as particular cases. In Figure 1 we give three examples where the (H, C, K)-coloring corresponds to the problems, k-Independent Set: does G have an independent set of size k?, k-Vertex Cover: does G have a vertex cover of size k?, and k-Almost Bipartite Graph does G contain a set of k vertices whose removal leaves a bipartite graph?. The shadowed regions represent the set C.

The complexity of the (H, C, K)-coloring problems was studied in [DST01, DST06+]. We proved a dichotomy for the list (H, C, K)-coloring problem: the problem is in P if $H - C$ is a bi-arc graph, otherwise it is NP-Complete. For the (H, C, K)-coloring, we proved that the problem is in P if $H - C$ is bipartite or contains a loop, and it is NP-Complete if $H - C$ is not a bi-arc graph, leaving



k-INDEPENDET SET k-VERTEX COVER k-ALMOST BIPARTITE

Fig. 1. The partially weighted graphs (H, C, K) for the problems k-INDEPENDENT SET k-VERTEX COVER and k-ALMOST BIPARTITE

open the dichotomy. In [DST04a], the interested reader can find a survey on results for different parameterization of H-coloring.

In this paper we are interested in finding a large class of (H, C, K) for which the (H, C, K)-coloring problems are *fixed parameter tractable*. Recall that the class FPT of fixed parameter tractable problems is formed by those parameterized problems that can be solved by an algorithm in time $O(f(k)n^c)$ where k is the parameter, n is the input size, and c is a constant independent of k. One of the fundamental tools for designing efficient fixed-parameter algorithms for decision problems is the so called *reduction to problem kernel*. Using this method we give a way to transform G, to a graph $\widehat{G}$, the *kernel*, such that:

- the transformation of G to $\widehat{G}$ can be done in time that is polynomial on the size of G,
- G has an list (H, C, K)-coloring if and only if $\widehat{G}$-has a list (H, C, K)-coloring,
- the size of $\widehat{G}$ depends only on the parameter k.

We isolate two special families, the *simple* and the *plain* partially weighted graphs (see definitions in section 2). For simple (H, C, K) we design linear time fixed parameter algorithms for the list (H, C, K)-problem. For the case of plain (H, C, K) we show that the (H, C, K)-coloring problem is in FPT. Notice that the definition of those families is an extension of the ones considered in [DST04b]. Let us mention that in [DST04b] we also provided a list of results for the counting version of the (H, C, K)-coloring and the list (H, C, K)-coloring.

In Figure 2 we summarise the parameters and notations used to give the time bounds of our algorithms. Although they will be defined in the next section we will use them directly in our theorems without restating.

The paper is organized as follows: in Section 2 we provide the basic definitions used in the paper.

In Section 3, we present fast exact algorithms for the list (H, C, K)-coloring problem for three basic graph families: (a) when either G is edge-less or H is a complete reflexive; (b) $H - C$ is edge-less; and (c) H is a complete bipartite graph.

In Section 4, we describe how to construct a kernel $\widehat{G}$, for a connected graph G. This construction allows us to solve the list (H, C, K)-coloring prob-

G	
g	nb of connected components
n	nb of vertices of G ($\|V(G)\|$)

(H, C, K)	
γ	nb of connected components f H
h	nb of vertices of H ($\|V(H)\|$)
c	nb of weighted vertices ($\|C\|$)
s	nb of non weighted vertices ($\|V(H) - C\|$)
k	total weight ($\sum_{c \in C} K(c)$)

Fig. 2. Notation and parameters used in the analysis of our parameterized algorithms

lem, in the case that (H, C, K) is plain, by running the exact algorithms obtained in Section 3 on the kernel.

In the case that G is not a connected graph, its connected components may be mapped in different connected components of H and more than one connected component of G can be mapped to the same component of H. To deal with such a case, in Section 5 we define a special structure called *signature*. A signature keeps information on all the ways in which a subset of the connected components of G can be mapped into components of (H, C, K) respecting the list L. Then we construct a special subset D of representatives of the connected components such that the size of D depends only on the parameter. We show that the existence of a non-empty signature for some subset of D is equivalent to the existence of a list (H, C, K)-coloring of G. Thus, checking for this property in linear time allow us to produce an efficient algorithm for the list (H, C, K)-coloring problem.

In Section 6, we present a parameterized reduction from the (H, C, K)-coloring problem for plain (H, C, K) to the list (H, C, K)-coloring for simple (H, C, K). The reduction is done through a series of sub-reductions that gradually transforms a plain (H, C, K) into a simpler (H, C, K).

In Section 7 we examine several extensions of our results. First, we present some negative results on the parameterized complexity of the (list) (H, C, K)-coloring problem. We identify some (H, C, K) where the corresponding problems are W[1]-hard. Recall that W[1] is a class of parameterized problems defined in [DF95a, DF95b] and the W[1]-hardness of a problem makes it quite unexpected that the problem belongs in FPT. Second, we modify our algorithms in order to construct a list (H, C, K)-coloring (if it exists) and we explain how the existing algorithms should be modified for different versions of the problem. Finally, we conclude in Section 8 with a discussion of our results and open problems.

2 Basic Definitions and Notation

Given a graph G, let $V(G)$ denote its vertex set and $E(G)$ denote its edge set. For $u, v \in E(G)$, $\{u, v\}$ denotes an edge between u and v and $\{u, u\}$ a loop at vertex u. We say that u and v are *adjacent in* G when $\{u, u\} \in E(G)$. If $U \subseteq V(G)$, $G[U]$ denotes the *subgraph of* G *induced by* U. Following the terminology in [FH98, FHH99], we say that a graph is *reflexive* if all its vertices have a loop, and that a graph is *irreflexive* when none of its vertices is looped. We denote by $g = g(G)$ the number of connected components of G. For $U \subseteq V(G)$, we use $G - U$ as a notation for the graph $G[V(G) - U]$. Define the neighborhood of a vertex v in G by $N_G(v) = \{u \in V(G) | \{u, v\} \in E(G)\}$. Given two graphs G and G', with no vertices in common, their *disjoint union* is the graph $G \cup G' = (V(G) \cup V(G'), E(G) \cup E(G'))$, and their *join* is the graph $G \oplus G' = (V(G) \cup V(G'), E(G) \cup E(G') \cup \{\{u, v\} \mid u \in V(G), v \in V(G')\}$.

Given a bipartite connected graph G, the vertex set $V(G)$ is splited among two sets $X(G)$ and $Y(G)$ so that every edge has an end point in the X-part and another in the Y-part. We assume that the X and Y parts of a connected bipartite graph are defined without ambiguity, for example setting the X part to be the part that contains the first vertex. Let $K_{r,s}$ denote the complete irreflexive bipartite graph on two parts with r and s vertices each. Observe that $K_{r,0}$ denotes a graph G with r vertices a no edges. Finally, K_n^r denotes the reflexive clique with n vertices.

For any function $\varphi : A \to B$ and any subset $A' \subseteq A$, we define the *restriction of* φ *to* A' by $\varphi|_{A'} = \{(a, b) \in \varphi \ \mid a \in D\}$. Given two functions $\varphi : A \to \mathbb{N}$ and $\psi : A' \to \mathbb{N}$, with $A' \cap A = \emptyset$, the *disjoint union* of φ and ψ, is a function from $(\varphi \cup \psi) : A \cup A' \to \mathbb{N}$ such that $(\varphi \cup \psi)(x) = \varphi(x)$ if $x \in A$, and $(\varphi \cup \psi)(x) = \psi(x)$ if $x \in A'$.

Given two functions $\varphi, \psi : A \to \mathbb{N}$, for any $x \in A$, we define $(\varphi + \psi)(x) = \varphi(x) + \psi(x)$. We say that $\phi \leq \psi$ if, for any $x \in A$, $\phi(x) \leq \psi(x)$. We say that $\varphi : A \to \mathbb{N}$ is positive if, for any $a \in A$, we have $\varphi(A) > 0$. We denote by $\varnothing$ the empty function.

We often need sets of consecutive indices. For $n, m \in \mathbb{N}$, we use the notation $[n] = \{1, \ldots, n\}$ and $[-m, n] = \{-m, \ldots, -1, 0, 1, \ldots, n\}$.

Given two graphs G and H, we say that a function $\chi : V(G) \to V(H)$ is an *H-coloring* of G, if for any edge $\{x, y\}$ of G, $\{\chi(x), \chi(y)\}$ is also an edge of H.

For a fixed graph H, given a graph G, an *(H, G)-list* is a function $L : V(G) \to 2^{V(H)}$ mapping any vertex of G to a subset of $V(H)$. For any $v \in V(G)$, the *list* of v is the set $L(v) \subseteq V(H)$. Given an (H, G)-list L, a *list H-coloring* of (G, L) (or a list H-coloring for short) is an H-coloring χ of G such that, for every $u \in V(G)$, $\chi(u) \in L(u)$.

A *partial weight assignment* on H is a pair (C, K), where $C \subseteq V(H)$ and $K : C \to \mathbb{N}$. We refer to the vertices of C as the *weighted vertices*. Given a partial weight assignment (C, K) on H, a mapping $\chi : V(G) \to V(H)$ is a *restrictive H-coloring* of G if χ is an H-coloring of G such that for all $a \in C$, we have $|\chi^{-1}(a)| = K(a)$.

For a fixed graph H, given an input graph G, the *H-coloring* problem asks whether there exists an H-coloring of G. In the same way, for a given graph G and an (H,G)-list L, the list *H-coloring* asks whether there exists a list H-coloring of (G,L). If in addition a partial weight assignment (C,K) is given and we ask for a restrictive or a list restrictive H-coloring, we get the *restrictive H-coloring* and the *restrictive list H-coloring* problems.

We consider the parameterization of the restrictive H-coloring problems in which the partial weight assignment (C,K) on H is selected as parameter. In order to simplify notation, we represent by a triple (H,C,K) the target graph and the partial weight assignment. We also extend the term restrictive to the set of colorings: An (H,C,K)-coloring of G is a restrictive H-coloring of G and (C,K), while a list (H,C,K)-coloring of G is a restrictive list H-coloring of G, an (H,G)-list L, and (C,K).

For a fixed (H,C,K), given an input graph G, the *(H,C,K)-coloring* problem asks whether there exists an (H,C,K)-coloring of G. The *list (H,C,K)-coloring* problem asks whether there exists an (H,C,K)-coloring of (G,L), when G and an (H,G)-list L are given as input.

Notice H-coloring and (H,C,K)-colorings behave differently for non connected graphs. An H-coloring of G is obtained as the union of the H-colorings of the connected components of G. However, the union of (H,C,K)-colorings for the connected components of G might not be an (H,C,K)-coloring of G. This is due to the fact that the total number of pre-images of the vertices in H might exceed the corresponding bound.

A partially weighted graph (H,C,K) is *positive* if the function K is positive. We say that (H,C,K) and (H',C',K') are *equivalent*, and we denote it as $(H,C,K) \sim (H',C',K')$, if, for any graph G, G has an (H,C,K)-coloring if and only if G has an (H',C',K')-coloring.

Next, we introduce some types of partially weighted graphs. An illustration of these types is given in Figure 3.

A partially weighted graph (H,C,K) with $C \neq \emptyset$, is a

- *1-component* if $E(H-C) = \emptyset$.
- *2-component* if H is a reflexive clique.
- *3-component* if H is a complete bipartite graph.
- *4-component* if $H[C]$ is a reflexive clique with all its vertices adjacent to one looped vertex x of $V(H) - C$.
- *6-component* if H is bipartite and
  - (1) if $C \cap X(H) \neq \emptyset$ and $C \cap Y(H) \neq \emptyset$, there are vertices $x \in X(H) - C$, $y \in Y(H) - C$, $a \in C \cap X(H)$, and $b \in C \cap X(H)$, such that $(x,y) \in E(H)$, x is adjacent to all the vertices in $C \cap Y(H)$, y is adjacent to all the vertices in $C \cap X(H)$, and $\{a,b\} \in E(H)$, or

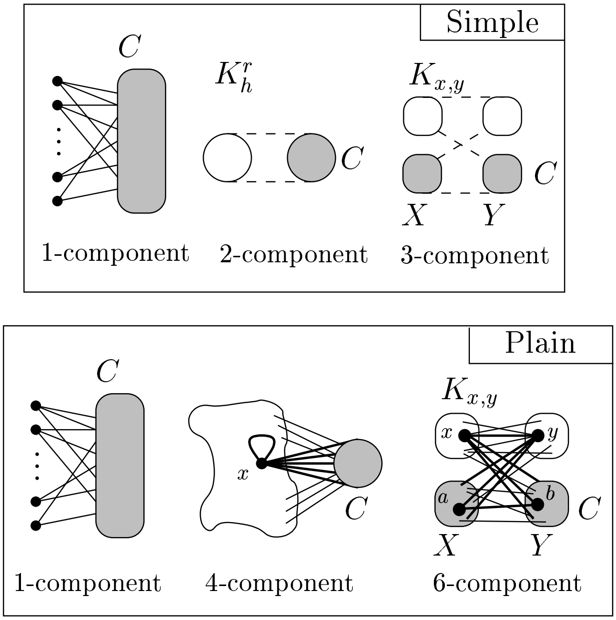


Fig. 3. Components for simple and plain (H, C, K)

– (2) if $C \cap X(H) \neq \emptyset$ but $C \cap Y(H) = \emptyset$, there are vertices $x \in X(H) - C$, $y \in Y(H)$, and $a \in C \cap X(H)$ such that $(x, y) \in E(H)$, and y is adjacent to all the vertices in $C \cap X(H)$, or
– (3) if $C \cap X(H) = \emptyset$ and $C \cap Y(H) \neq \emptyset$, there are vertices $x \in X(H) - C$, $y \in Y(H) - C$, and $b \in C \cap X(H)$, such that $(x, y) \in E(H)$ and x is adjacent to all the vertices in $C \cap Y(H)$.

Notice that the cases of 2 and 3-components are the only ones in which the graph H is required to be connected. Furthermore the connected components of any other type of component must be of the same type.

A partially weighted graph (H, C, K) with $C = \emptyset$, is a

- *1-component* when H is edge-less,
- *2-component* when H is a reflexive clique,
- *3-component* when H is a complete bipartite graph,
- *4-component* when H has a looped vertex,
- *6-component* when H is bipartite.

(H, C, K) is said to be *simple* when, for each connected component H' of H, $(H', C \cap H', K|_{V(H')})$ is either a 1-component, a 2-component, or a 3-component.

(H, C, K) is said to be *plain* when, for each connected component H' of H, $(H', C \cap H', K|_{V(H')})$ is either a 1-component, a 4-component, or a 6-component.

We assume that (H, C, K) is pre-processed once in our algorithms. This preprocessing consists on the computation of the connected components of H, and their classification according to the above types. The overall additional computation can be performed in $O(h^2)$ steps. In Sections 3, 4, and 5, we give algorithms for the list (H, C, K)-problem assuming that the input G is a graph without loops. This does not harm the generality of our results as looped vertices of G should be mapped to looped vertices of H. Therefore, an equivalent instance can be constructed if, for any looped vertex v in G, we replace $L(v)$ by $L(v) \cap \{a \mid a \text{ is a looped vertex in } H\}$, thus removing all loops in G.

All through the paper, for a given (H, C, K), we use the notation $S = V(H) - C$, $h = |V(H)|$, $c = |C|$, $s = |S|$ and $k = \sum_{c \in C} K(c)$.

3 Exact Algorithms for List (H, C, K)-coloring in Particular Cases

The fixed parameter algorithms for deciding the existence of a list (H, C, K)-coloring problem, when (H, C, K) is a 1, 2 or 3-component, are based on the kernelization technique. We first construct a kernel, and then we solve the (H, C, K)-coloring problem on it. In this section, we devise fast algorithms for particular cases of H, G and $H - C$.

3.1 G is Edge-Less or H is a Reflexive Clique

For the first two cases ($E(G) = \emptyset$ or $H = K_h^r$) we consider the following algorithm:

```
function Basic(H, G, L)
    begin
        for all v ∈ V(G) do
            if L(v) = ∅ then return false  end if
        end for
        return true
    end
```

Theorem 3.1. *Let H and G be two graphs such that either G has no edges or $H = K_h^r$. Given a (H, G)-list L, function* Basic *correctly decides in $O(hn)$ steps whether a list H-coloring of (G, L) exists.*

Proof. Notice that, if G is edge-less or H is a reflexive clique, any mapping from $V(G)$ to H is a valid H-coloring of G. Thus, given a (H, G)-list L, there is a list H-coloring of (G, L) if and only if for all $v \in V(G)$, $L(v) \neq \emptyset$. □

Our next results shows how to relate list (H, C, K)-colorings of (G, L) with maximum weight matchings in an associated bipartite graph. Let (H, C, K) be a partially weighted graph, and let L be an (H, G)-list. We define the *associated bipartite* weighted graph $B = B(H, C, K, G, L)$ where

- $X(B) = V(G)$,
- for any $a \in C$ let $A(a)$ be a set with $K(a)$ new vertices,
- $Y(B) = \bigcup_{a \in C} A(a)$,
- for $u \in X(B)$ and $b \in Y(B)$, there is an edge $\{u, b\}$ if, for some $a \in L(u) \cap C$, $b \in A(a)$.

Finally, if $v \in X(B)$ and $L(v) \cap S \neq \emptyset$, the weight of all edges having v as endpoint is 1. All the remaining edges get weight k. We usually refer to $X(B)$ as the G-part and to $Y(B)$ as the C-part of B. In Figure 4 we give an example of the construction, notice that the graph B depends on the structure of L but not on the edges present in either G or H.

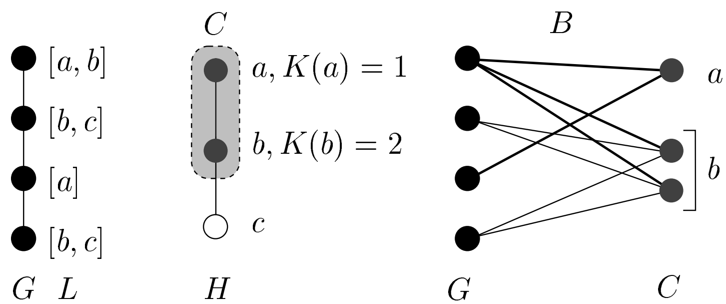


Fig. 4. A partially weighted graph (H, C, K), a graph G, a (G, H)-list, and the associated weighted bipartite graph $B = B(H, C, K, G, L)$. In B, the "fat" edges have weight 3 and the rest have weight 1

Lemma 3.2. *Let (H, C, K) be a partially weighted graph, G a graph, and L an (H, G)-list. The weighted bipartite graph $B = B(H, C, K, G, L)$ can be computed in $O((k + h)n)$ steps.*

Proof. For any vertex in $V(G)$, its list of neighbors can be computed in $h + k$ steps by examining the bits in $L(u)$ and adding the corresponding elements. The list for the other half can be obtained by reversing the previously computed information in the same time bounds. ◻

Next we relate the existence of list colorings to the weight of a maximum matching in B.

Lemma 3.3. *Let (H, C, K) be a partially weighted graph, G a graph, and L an (H, G)-list and let $B = B(H, C, K, G, L)$ be the associated bipartite graph.*

Assume that $E(G) = \emptyset$ or that $H = K_h^r$. Let $p = |\{v \in V(G) \mid L(v) \subseteq C\}|$. If $p > k$, there is no list (H, C, K)-coloring of (G, L). Otherwise, there is a list (H, C, K)-coloring of (G, L) if and only if the weight of a maximum matching in B is $kp + k - p$.

Proof. Note that the maximum number of vertices in G that can be mapped to C is k. Furthermore, all the vertices $v \in V(G)$ with $L(v) \subseteq C$ must be mapped to C. Therefore, if $p > k$, (G, L) does not have a list (H, C, K)-coloring.

Assume now that $p \leq k$ and let us prove the second part of the lemma.

Assume that χ is a list (H, C, K)-coloring of (G, L). Let us construct a maximum weight matching for B. For every $a \in C$, assign the vertices in $\chi^{-1}(a)$ to the different copies of a in B. This assignment provides a matching with weight $pk+k-p$. Observe that, by construction, $pk+k-p$ is the maximum weight of any selection of k vertices in the G-part of B, as the number of edges with weight k is maximized. Therefore, the constructed matching has maximum weight.

For the reverse implication note that, by definition of B, in any maximum weight matching M of B with $w(M) = kp + k - p$, all the vertices in the C-part are matched. Notice that the set of edges with weight k in the matching is maximum. Furthermore, the matching includes all the vertices in G that must be mapped to C. Therefore, the matching M provides an assignment of vertices of G to vertices of $Y(B)$ composing this mapping with a the function φ such that, for $y \in A(a)$, $\varphi(y) = a$ we get a valid (H, C, K)-coloring of (G, L). ◻

Consider the following algorithm:

```
function List-basic(H, C, K, G, L)
    begin
        p = 0
        for all v ∈ V(G) with L(v) ∩ S = ∅ do
            p = p + 1
            if p > k then return ∅ end if
        end for
        Construct the weighted bipartite graph B(H, C, K, G, L)
        Compute a maximum matching M of B
        if w(M) < kp + k − p then return false  end if
        V' = {v ∈ V(G) | v is not matched in M}
        G' = (V', ∅)
        forall v ∈ V' set L'(v) = L(v) ∩ S end for
        return Basic(H[S], G', L')
    end
```

Theorem 3.4. *Let (H, C, K) be a partially weighted graph, let G be a graph, and let L be a (H, G)-list. When G has no edges or $H = K_h^r$, function List-basic correctly decides in $O((k+h)n + nk \log k\sqrt{n+k})$ steps whether a list (H, C, K)-coloring of (G, L) exists.*

Proof. The correctness follows from Lemma 3.3. From Lemma 3.2 we have that computing the associated bipartite graph takes $O((k+h)n)$ time. This bound dominates all the remaining steps except from computing the maximum weighted matching in B. We use the Gabow-Tarjan algorithm in [GT89] for computing a maximum weighted matching of a bipartite graph in time $O(m\sqrt{n} \log N)$, where m is the number of edges, n the number of vertices and N is the maximum weight of an edge. As B has at most nk edges and $n + k$ vertices, the claimed bound follows. ◻

3.2 $H - C$ is Edge-less

We use the self-reducibility of the list H-coloring problem to design an algorithm for the list (H, C, K)-coloring when $E(H - C) = \emptyset$. Recall that the self-reducibility can be used in the following well known way: Consider a graph G, a partially weighted graph (H, C, K), an (H, G)-list L, a vertex $v \in V(G)$, and a vertex $a \in C$ with $K(a) > 0$. Define $G' = G[V(G) - \{v\}]$, $K'(b) = K(b)$ for $b \neq a$ and $K'(a) = K(a) - 1$. For any $u \in V(G')$ let

$$L'(u) = \begin{cases} L(u) & \text{if } u \notin N_G(v), \\ L(u) \cap N_H(a) & \text{otherwise.} \end{cases}$$

Then (G, L) has a list (H, C, K)-coloring χ where $\chi(v) = a$ if and only if (G', L') has a list (H, C, K')-coloring.

The idea behind the algorithm is to reduce the computing time by using a truncated search tree. Each edge in G must have at least one of its endpoints mapped to C, the algorithm tries both possibilities recursively. Each recursive call removes one vertex v that has already being mapped to some $c \in C$, modifies the lists of v's neighbors to guarantee that the recursive call is correct (according to self-reducibility), and decreases $K(c)$ by one. The recursion finishes either when the remaining weight is zero or when G looses all its edges.

```
function Self-red(H, C, K, G, L)
  begin
    if L(v) = ∅ for some v ∈ V(G) then return false end if
    if K = 0 ∧ E(G) ≠ ∅ then return false end if
    if K = 0 ∧ E(G) = ∅ then return Basic(H[S], G, L) end if
    if E(G) = ∅ then return List-basic(H, C, K, G, L) end if
    Select an edge e ∈ E(G)
    for both endpoints v of e do
    G' = G − {v}
```

```
        for all a ∈ C with K(a) > 0 do
            for all b ∈ C − {a} set K'(b) = K(b) end for
            K'(a) = K(a) − 1
            for all u ∈ V(G') set L'(u) = L(u) end for
            for all u ∈ N_G(v) set L'(u) = L'(u) ∩ N_H(a) end for
            if Self-red(H, C, K', G', L') then return true end if
        end for
    end for
    return false
end
```

Theorem 3.5. *Let (H, C, K) be a partially weighted graph, where $H - C$ is edge-less. Given an input graph G, function* Self-red *decides whether there is a list (H, C, K)-coloring of (G, L) in $O\left(2^k c^k \left((k+h)n + nk\sqrt{n+k}\log k\right)\right)$ steps.*

Proof. The correctness of Self-red follows from the self-reducibility defined at the beginning of this subsections, and from the fact that, for any edge in G, a (H, C, K)-coloring of G should map at least one of its endpoints to C. The algorithm applies the self-reduction on any possible mapping of any endpoint of a selection of $\min\{k, |E(G)|\}$ edges to some of the vertices in C. If for at least one of these choices, (G', L') has a list (H, C, K')-coloring that can be extended to a list (H, C, K)-coloring of (G, L) we compute one. Otherwise G does not have any. There are two trivial base cases, the case $L(v) = \emptyset$ or the case when $K = \mathbf{0}$ and $E(G) \neq \emptyset$. In both cases there are no colorings of (G', L'). In the other two base cases $K = \mathbf{0}$ or $E(G) = \emptyset$, we are in one of the basic cases that can be solved according to Corollaries 3.1 and 3.4 using functions Basic or List-basic respectively.

The time required for the application of the self reduction involved in the main loop is $O(n + k)$. Furthermore, we consider at most k edges and the 2^k possibilities of end-point selections, together with the c^k possible assignments for any selection. Thus the overall number of calls to the extreme cases is $O(2^k c^k)$. Taking into account the time bound for the List-basic function given in Corollary 3.4, the claimed bound follows. □

3.3 *H* is a Complete Bipartite Graph and *G* is Connected

To complete the picture, we give the steps needed to solve the list (H, C, K)-coloring for the case when H is a complete bipartite graph and G is connected.

Consider the following algorithm.

```
function Bipartite(H, C, K, G, L)
    begin
        for all u ∈ X(G) do
            Set L_1(u) = L(u) ∩ X(H)
```

```
        Set L_2(u) = L(u) ∩ Y(H)
    end for
    for all u ∈ Y(G) do
        Set L_1(u) = L(u) ∩ Y(H)
        Set L_2(u) = L(u) ∩ X(H)
    end for
    if (List-basic(H[X(H)], C, K, G[X(G)], L_1)
        ∧ List-basic(H[Y(H)], C, K, G[Y(G)], L_1)
        then return true
        else
            return (List-basic(H[Y(H)], G[Y(G)], L_2)
             ∧ List-basic(H[X(H)], C, K, G[X(G)], L_2))

        end
```

Theorem 3.6. *Let (H, C, K) be a partially weighted graph where $H = K_{x,y}$. Given G a connected bipartite graph and an (H, G)-list L, function* Bipartite *correctly decides in $O((k + h)n + nk \log k\sqrt{n + k})$ steps whether a list (H, C, K)-coloring of (G, L) exists.*

Proof. Given a connected bipartite graph G together with a (H, G)-list L, we define two (H, G)-lists, L_1 and L_2, as follows: for any $u \in V(G)$

$$L_1(u) = \begin{cases} L(u) \cap X(H) \text{ if } u \in X(G), \\ L(u) \cap Y(H) \text{ if } u \in Y(G), \end{cases} \quad L_2(u) = \begin{cases} L(u) \cap Y(H) \text{ if } u \in X(G), \\ L(u) \cap X(H) \text{ if } u \in Y(G). \end{cases}$$

Taking into account that in any coloring the vertices in one part (X or Y) of G must be mapped to only one part in H and that $H = K_{x,y}$, we have that there is a list (H, C, K)-colorings of (G, L) if and only if there is a list $(H[X(H)], C, K)$-coloring of $(G[X(G)], L_1)$ and a list $(H[Y(H)], C, K)$-coloring of $(G[Y(G)], L_1)$, or there is a list $(H[X(H)], C, K)$-coloring of $(G[Y(G)], L_2)$ and a list $(H[Y(H)], C, K)$-coloring of $(G[X(G)], L_2)$. Thus the algorithm is correct.

The time bound follows from Corollaries 3.1 and 3.4. □

4 The List (H, C, K)-coloring Problem for Connected Graphs

We start presenting a generic algorithm, called List-Coloring, for the case where G is a connected graph. In the following subsections we show the adjustments needed for the different types of components.

Our first step is to introduce an equivalence relation on the vertices of a connected graph G. This equivalence will be used as the main tool to construct the kernels. Assume that (H, C, K), a graph G, and an (H, G)-list L are given. We define $\mathcal{P}$ to be the partition of $V(G)$ induced by the equivalence relation,

$$v \sim u \text{ iff } (N_G(v) = N_G(u) \;\wedge\; L(v) = L(u))\,.$$

For $v \in V(G)$, P_v denotes the set $\{u \mid u \sim v\}$. We say that $R \subseteq V(G)$ is a *closed set* of the partition $\mathcal{P}$, if for any $v \in R$ we have $P_v \subseteq R$.

For $v \in V(G)$ let P_v^k be P_v if $|P_v| \leq k$, otherwise P_v^k is a subset of P_v with $k+1$ elements. Let R be a closed set, the *graph $\widehat{G}$ associated to* (G, R) is the graph induced by all the vertices in R, and for the classes P_v with $v \notin R$, all the vertices in P_v^k.

Note, that to ensure that $\widehat{G}$ has bounded size it is required that $\mathcal{P}$ has a small number of classes. Furthermore, $V(\widehat{G})$ keeps all the vertices of some classes with few vertices (the set R) and a restricted number of representatives for the remaining classes.

Next lemma shows the property which suffices for showing that the graph associated to a closed set is an adequate kernel, provided that its size is small.

Lemma 4.1. *Let (H, C, K) be a partially weighted graph. Given a connected graph G together with an (H, G)-list L and a closed set R, let $\widehat{G}$ be the graph associated to (G, R). Then, (G, L) has a list (H, C, K)-coloring if and only if $(\widehat{G}, L)$ has a list (H, C, K)-coloring.*

Proof. Let w be an (H, C, K)-coloring of $(\widehat{G}, L)$. We extend the coloring w to a coloring χ of G as follows: For those vertices $u \in V(G)$, we set $\chi(u) = w(u)$. For those vertices $u \in V(G) \setminus V(\widehat{G})$, we know that $|P_u| > k+1$, therefore there are exactly $k+1$ vertices in $V(\widehat{G})$ from P_u, which implies that there exists $v \in P_u$ for which $w(v) \in S$. Set $\chi(u) = w(v)$. The definition of $\mathcal{P}$ ensures that the obtained coloring is an (H, C, K)-coloring of (G, L).

To prove the reverse implication, let χ be a (H, C, K)-coloring of (G, L), we construct a coloring w of $\widehat{G}$. Notice that for any equivalence class Q with $|Q| > k+1$, we have that $|\chi(Q) \cap S| > 1$, therefore one vertex $a_Q \in \chi(Q) \cap S$ can be selected. For any $d \in C$ let $d_Q = |\chi^{-1}(c) \cap Q|$. Define $w : V(\widehat{G}) \to V(H)$ as follows: For $u \in R$, set $w(u) = \chi(u)$; otherwise, if $u \notin R$ but $|P_u| \leq k+1$, set $w(u) = \chi(u)$. To each representative of a class Q with $|Q| > k+1$, assign in a round robin way, vertex $d \in C$ to d_Q vertices, and complete the coloring assigning a_Q to the remaining vertices in $Q \cap V(\widehat{G})$. Again the definition of $\mathcal{P}$ guarantees that w is a (H, C, K)-coloring of $(\widehat{G}, L)$. □

The generic algorithm for deciding the existence of a list (H, C, K)-coloring problem uses $\widehat{G}$ as kernel:

function List-Coloring(H, C, K, G, L)
  input A partially weighted graph (H, C, K).
    A connected graph G and a (H, G)-list L.
  begin
    $k = \sum_{a \in C} K(a); s = |V(H) - C|;$
    $R =$ Closed-set(G, k)

```
        if R[0] = 0 then return NO end if
        Ĝ = Kernel(G, R, k + 1)
        return Exact(H, C, K, Ĝ, L)
    end
```

List-Coloring(H, C, K, G, L) computes first the adequate closed set, defined in a different way for each type of component. The adequate definition of R and the definition of $\mathcal{P}$ will guarantee that $\widehat{G}$ has bounded size. In a second step, the algorithm checks whether there is a list (H, C, K)-coloring of $(\widehat{G}, L)$, using the exact algorithms presented in the previous section. The central **if** detects whether the non-existence of list (H, C, K)-colorings of G has been observed while computing the corresponding closed set.

In the following subsections we give the definition of the closed-set that guarantees that $\widehat{G}$ is a kernel for each type of component and fix the remaining components for the analysis of the adaptation of algorithm List-Coloring.

4.1 The Case of 1-components

To obtain the adequate kernelization, we start considering a classical kernelization for vertex cover due to Buss and Goldsmith [BG93].

Given a graph G and an integer k, the *k-splitting* of G is the partition of $V(G)$ in three sets (R_1, R_2, R_3), where R_1 is the set of vertices in G of greater than k, R_2 is formed by the non isolated vertices in $G' = G[V(G) - R_1]$, and R_3 contains the isolated vertices in G'.

Lemma 4.2. *Let (H, C, K) be a partially weighted graph, where $H - C$ is edge-less. Given a connected graph G, let (R_1, R_2, R_3) be the k-splitting of G. Then $R_1 \cup R_2$ is a closed set. Furthermore, if $|R_1| > k$ or $|R_2| > k^2 + k$, then there are no (H, C, K)-colorings of (G, L).*

Proof. Let us prove the first claim. Notice that the number of neighbors of a vertex in R_1 is bigger than k and thus any vertex with the same neighborhood must belong to R_1. The vertices in R_2 are the only vertices that have at most k neighbors, all of them in $R_1 \cup R_2$ and at least one in R_2. Therefore $R_1 \cup R_2$ is a closed set.

For the second claim. Notice that C is a vertex cover in H. Therefore, for any (H, C, K)-coloring χ of G, $\chi^{-1}(C)$ must be a vertex cover in G of size $k = \sum_{i \in C} k_i$. Therefore, for every $v \in \chi^{-1}(V(H) - C)$, $d_G(v) \leq k$. In any (H, C, K)-coloring of G, if there exists a $v \in G$ such that $d_G(v) > k$, then v must be mapped to C, so $|R_1| \leq k$. Furthermore, at most k vertices outside of R_1 can be mapped to C. The vertices in R_2 not mapped to C must be connected to at least one vertex in R_2 mapped to C. Thus, taking into account that, for all $v \in R_2$, $d_G(v) \leq k$, we have that $|R_2| \leq k^2 + k$. □

Lemma 4.3. *Let (H, C, K) be a partially weighted graph, where $H - C$ is edge-less. Given a connected graph G and an (H, G)-list, let (R_1, R_2, R_3) be* *the k-splitting of G and let $\widehat{G}$ be the graph associated to $(G, R_1 \cup R_2)$. If $|R_1| \leq k$ and $|R_2| \leq k^2 + k$, then $\widehat{G}$ has at most $k^2 + 2k + (k+1)2^{k+h}$ vertices. Furthermore, $\widehat{G}$ can be obtained in time $O((k+h)n + 2^{k+h})$.*

Proof. The following algorithm constructs the closed set associated to the k-splitting of G.

```
function Closed-set(G, k)
    begin
        ν = 0
        for all v ∈ V(G) do
            if d_G(v) > k then R[v] = 1; ν = ν + 1 else R[v] = 2 end if
        end for
        if ν > k then R[0] = 0; return R else R[0] = ν; ν = n − ν end if
        for all v with R[v] = 2 ∧ N_G(v) ⊆ R_1 do
            R[v] = 3; ν = ν − 1
        end for
        if ν > k^2 + k then R[0] = 0 end if
        return R
    end
```

From the above description, taking into account that G is represented by an array of linked list of sorted neighbors, we have that function Closed-set finishes in $O(kn)$ steps. Furthermore, if $|R_1| \leq k$ and $|R_2| \leq k^2+k$, it computes the closed set associated to the k-splitting. Notice that the first iteration separates the vertices in R_1 from the remaining vertices, by assigning to them label 1 or 2. Once this is done, the second iteration identifies those vertices with label 2 whose neighborhood is contained in R_1. Therefore computing R_2 (vertices with label 2) and R_3 (vertices with label 3).

By definition $\widehat{G}$ contains the vertices in $R_1 \cup R_2$, at most $k^2 + k$, and at most $k + 1$ vertices from each class. As the vertices in R_3 are connected only to vertices in R_1 the number of different neighborhoods is at most 2^k. Thus taking into account that the number of possible lists is at most 2^h we get that $|V(\widehat{G})| \leq k^2 + k + (k+1)2^{k+h}$.

Notice that, by construction, the vertices in R_3 are connected only to vertices in R_1. To obtain $\widehat{G}$ we select a set of vertices U from R_3. The set U is initially empty. We use an auxiliary rooted tree T whose nodes hold a counter. Each node in the tree identifies a pair formed by a subset of R_1, defined by a sorted list of vertices, and a subset of $V(H)$, identified also by a sorted list. Notice that such a tree has size bounded by 2^{k+h}.

We process the vertices in R one after the other. Let u be a vertex in R_3. We check whether the counter associated to the pair $(N_G(u), L(u))$ has not reached $k+1$. If so we increase the counter by one and keep u in U, otherwise we discard u.

Once the set U is computed, we set $\widehat{G} = G[U \cup R]$. Notice that it takes $k + h$ steps to identify the associated pair. Furthermore, each vertex in $\widehat{G}$

has degree at most k thus $O(kn)$ additional steps are needed to construct the kernel. By the results obtained in the previous lemma, the total time is $O\left((k+h)n+2^{k+h}\right)$. ◻

Finally, we put together the two pieces to obtain an algorithm for the case of 1-components.

Theorem 4.4. *Let (H,C,K) be a partial weight assignment, where $E(H-C)=\emptyset$. Given an input graph G and a (H,G)-list L, there is an algorithm that decides whether there is a list (H,C,K)-coloring of (G,L) in time*

$$O\left((h+k)n+2^{k+h}+2^{5k/2}c^k p(k,h)\right),$$

for some polynomial p.

Proof. The correctness of the implementation of algorithm List-coloring follows from Lemmata 4.1, 4.2, and 4.3 together with Theorem 3.5.

Recall that, according to Lemma 4.3, $|V(\widehat{G})| \leq k^2+2k+(k+1)2^k = O(k^2+k2^k)$ and constructing $\widehat{G}$ takes $O((h+k)n+2^{k+h}$ steps. The correct exact algorithm to solve the kernel problem is the function Self-red, described in page 383. This algorithm requires

$$O\left(2^k c^k \left((k+h)|V(\widehat{G})| + |V(\widehat{G})| \cdot k\sqrt{|V(\widehat{G})|+k}\log k\right)\right)$$

which gives a bound of

$$O\left((h+k)n+2^{k+h}+2^k c^k((k+h)(k^2+k2^k)+(k^2+k2^k)^{\frac{3}{2}}k^2)\right).$$

◻

In Theorem 4.4, we may assume that both G and H are not necessarily connected. We can do this as we never required H to be connected in this subsection. This observation will be useful in Section 5 where we consider all connected 1-components of H as a unique unified 1-component.

4.2 The Kernel for 2-components

We start with an easy and trivial observation: for any partially weighted graph (H,C,K) with $H=K_h^r$, there exists a list (H,C,K)-coloring of (G,L) if an only if there is a list (H,C,K)-coloring of (G',L), where $G'=(V(G),\emptyset)$. Without loss of generality, in the remaining of this section, we assume that $E(G)=\emptyset$. This fact simplifies the structure of the equivalence classes in relation $\mathcal{P}$. Notice that the only existing neighborhood is the empty set. Therefore, we have to take into account only the list associated to each vertex.

Given a graph G together with an (H,G)-list L, define a *list splitting* of (G,L) as a partition of $V(G)$ in two sets, (N_1,N_2), where $N_1=\{v\in V(G) \mid L(v)\subseteq C\}$.

The following condition follows from the previous definition.

Lemma 4.5. *Let (H,C,K) be a weight assignment, where $H = K_h^r$. Given a graph G and an (H,G)-list L, let (N_1,N_2) be the list splitting of (G,L). Then N_1 is a closed set. Furthermore, if $|N_1| > k$, there are no list (H,C,K)-colorings of (G,L).*

Our next result shows that we can use the graph associated to N_1 as kernel.

Lemma 4.6. *Let (H,C,K) be a partially weighted graph with $H = K_h^r$. Given a graph G and a (H,G)-list L, let (N_1,N_2) be the list-splitting of (G,L), and let $\widehat{G}$ be the graph associated to (G,N_1). If $|N_1| \leq k$, then $\widehat{G}$ has at most $k+(k+1)2^h$ vertices. Furthermore, $\widehat{G}$ can be obtained in time $O(hn+kh2^h)$.*

Proof. In a way similar to the one used in the case of 1-components, the closed set associated to the list-splitting can be computed, using the following $O(hn)$-algorithm:

```
function Closed-set(G, k)
    begin
        ν = 0
        for all v ∈ V(G) do
            if L(v) ⊆ C then R[v] = 1; ν = ν + 1 else R[v] = 0 end if
        end for
        if ν > k then R[0] = 0; return R else R[0] = 1 end if
    end
```

Note that the case $|N_1| > k$ is recorded by setting $R[0] = 0$.

As G has no edges, the number of equivalence classes is at most 2^h. Therefore, each class can be identified by a string of length h representing a subset of H. Keeping an array we can process all the vertices in N_2 in a way similar to that used in the proof of Lemma 4.3. For every vertex $v \in N_2$, we check whether the entry associated to $L(v)$ has reached the value $k+1$. If so we discard the vertex, otherwise we keep the vertex in U and increment the counter. Finally, $\widehat{G}$ is the graph induced by $N_1 \cup U$.

The number of steps needed per vertex is at most h, therefore the overall computation of the kernel takes time $O(hn + h(k + (k+1)2^h))$ steps. □

Putting together the previous results, following the schema given by algorithm List-coloring, and taking into account that the algorithm List-basic, given in Page 382 solves the kernel problem, from Lemmata 4.1, 4.5, and 4.6, and Theorem 3.4 we have

Theorem 4.7. *Let (H,C,K) be a partially weighted graph with $H = K_h^r$ and let G be a connected graph, there is an algorithm that decides whether there is a list (H,C,K)-coloring of (G,L) in time*

$$O(hn + k^{5/2}2^{3h/2}\log k)$$

4.3 The Kernel for 3-components

Observe that for any partially weighted graph (H, C, K) with $H = K_{x,y}$, and given a (H,G)-list L, (G, L) has a list (H, C, K)-coloring if and only if (G', L) has a list (H, C, K)-coloring, where $G' = (X(G), Y(G), X(G) \times Y(G))$. Therefore we may assume that G is a complete bipartite graph.

Let (H, C, K) be a partially weighted graph with $H = K_{x,y}$. Given a complete bipartite graph G together with a (H, G)-list L, the *bipartite splitting* of (G, L) with respect to (H, C, K) is a partition of $V(G)$ in three sets, (M_1, M_2, M_3) where $M_3 = \{v \in V(G) \mid L(v) \not\subseteq C\}$, $M_1 = X(G) - M_3$ and $M_2 = Y(G) - M_3$.

Note that all the vertices in M_1 and M_2 must be mapped into C. Therefore, we obtain the following result,

Lemma 4.8. *Let (H, C, K) be a partially weighted graph with $H = K_{x,y}$. Given a complete bipartite graph G, let (M_1, M_2, M_3) be the bipartite splitting of (G, L). Then, $M_1 \cup M_2$ is a closed set. Furthermore, if $|M_1| + |M_2| > k$, there are no list (H, C, K)-colorings of (G, L).*

The next result gives the kernel conditions for 3-components.

Lemma 4.9. *Let (H, C, K) be a partially weighted graph with $H = K_{x,y}$. Given a connected bipartite graph G and an (H,G)-list L, let $M = (M_1, M_2, M_3)$ be the bipartite splitting of (G, L) and let $\widehat{G}$ be the graph associated to $(G, M_1 \cup M_2)$. If $|M_1| + |M_2| \leq k$, $\widehat{G}$ has at most $k + (k + 1)2^{h+1}$ vertices. Furthermore, $\widehat{G}$ can be obtained in time $O(hn + kh2^h)$.*

Proof. Working in a way similar to that of the case of 1 and 2-components, we devise first an algorithm for computing the associated closed set. The following function constructs the bipartite splitting of (G, L). At the end of its execution $R[0]$ will hold a 0 when $|M_1| + |M_2| > k$; and a 1 otherwise.

```
function Closed-set(G, k)
    begin
        ν1 = 0
        for all v ∈ X(G) do
            if L(v) ⊆ C then R[v] = 1; ν1 = ν1 + 1 else R[v] = 3 end if
        end for
        ν2 = 0;
        for all v ∈ Y(G) do
            if L(v) ⊆ C then R[v] = 2; ν1 = ν1 + 1 else R[v] = 3 end if
        end for
        if ν1 + ν2 > k
           then R[0] = 0
           else R[0] = 1 end if
        end if
        return R
    end
```

As G is a complete bipartite graph, the number of equivalence classes is at most 2^{h+1}. Furthermore, each class can be identified by a string of length h representing a subset of H and a letter x or y representing whether the neighborhood of the vertex is $X(G)$ or $Y(G)$. Using an array as data structure, we can process all the vertices that do not belong to the closed set, in a way similar to the one used in the proof of Lemma 4.3. For every vertex v, we check whether the entry associated to $L(v)$ and their neighborhood have reached the value $k+1$. In the positive case, we discard the vertex, otherwise we keep the vertex in a set U and increment the counter. Finally, $\widehat{G}$ is the graph induced by $M_1 \cup M_2 \cup U$.

The number of steps needed per vertex is at most $h+1$, therefore the overall computation of the kernel takes $O(hn + hk2^h)$ steps. ☐

Putting together the previous results, following the schema given by algorithm List-coloring, and taking into account that algorithm Bipartite solves the kernel problem, from Lemmata 4.1, 4.8, and 4.9, together with Theorem 3.6 we have

Theorem 4.10. *Let (H, C, K) be a partially weighted graph with $H = K_{x,y}$ and let G be a connected bipartite graph. There is an algorithm that decides whether exists a list (H, C, K)-coloring of (G, L), which works in time*

$$O(hn + hk2^h + k^{5/2}2^{3h/2} \log k)$$

5 The List (H, C, K)-coloring Problem for Simple Partially Weighted Graphs

When dealing with simple partially weighted graphs, those formed by 1, 2, or 3-components, we have to take care of the connected components of G. The key idea is that, when checking for the existence of a list (H, C, K)-coloring, it is enough to keep information about whether a particular component of (H, C, K) can be partially filled by a legal list coloring of a component G_i of G. Furthermore, as H has bounded size, we can classify the components of G and maintain a small set of representatives.

Assume that G has g connected components $G_1, \ldots, G_g$. Our purpose is to define a *signature* associated to subsets of connected components of G, formally subsets of the set $[g]$.

Assume that the number of 3-components in H is ρ, and the number of 2-components is η. A 3-component H_ι of H has a negative index ($\iota \in [-\rho, -1]$) and a 2-component H_ι has a positive index ($\iota \in [1, \eta]$). The union of the remaining 1-components of H is considered as a unique 1-component H_0 (not necessarily connected).

Before defining formally signatures, let us introduce some additional notation. For any $\iota \in [-\rho, \eta]$, define $C_\iota = C \cap V(H_\iota)$, $K_\iota = K|_{V(C_\iota)}$, and $\mathcal{W}_\iota = \{W : C_\iota \to \mathbb{N} \mid \mathbf{0} \leq W \leq K_\iota\}$. Let $k_\iota = sum_{a \in C_\iota} K_\iota(a)$,

$k_\iota^* = \prod_{a \in C_\iota} K_\iota(a)$ and $k^* = \prod_{a \in C} K(a)$. Observe that $k^* = \prod_{\iota \in [-\rho,\eta]} k_\iota^*$ and that $|\mathcal{W}_\iota| = k_\iota^*$. Finally, define $\mathcal{W}_\iota^+ = \mathcal{W}_\iota - \{\mathbf{0}\}$.

Given a graph G and an (H,G)-list G, for $i \in [g]$ and $\iota \in [-\rho,\eta]$, let $L_{i\iota}$ be the (H_ι, G_i)-list obtained by restricting L to G_i and H_ι. Given $i \in [g]$, $\iota \in [-\rho,\eta]$, we use $\widehat{G}_{i\iota}$ to denote the kernel obtained from G_i and H_ι according to the type of H_ι and following the definitions given in the previous section.

Our next definition is the basis for classifying the components of G according to the combinations of components of H and partial weight assignments with correct list colorings. Given $i \in [g]$, $\iota \in [-\rho,\eta]$, and $W \in \mathcal{W}_\iota$, define $\alpha(i,\iota,W) = 1$ or 0 depending on whether there is or is not a list-(H_ι, C_ι, W)-coloring of $(\widehat{G}_{i\iota}, L_{\iota i})$.

We isolate the information that must be captured from a set of components to certify the existence of a list coloring. Given $A \subseteq [g]$, and $\iota \in [-\rho,\eta]$ define:

$$\begin{aligned}\mathcal{F}(A,\iota) = \{ \ & (Z,f) \mid Z \subseteq A, f : Z \to \mathcal{W}_\iota^+, \\ & \sum_{i \in Z} f(i) = K_\iota \wedge \forall i \in Z \ \alpha(i,\iota,f(i)) = 1\}\end{aligned}$$

The elements of $\mathcal{F}(A,\iota)$ are pairs (Z,f) where formed by a selection of components of G and a selection of weights with the guarantee that there are list colorings that allow us to fill completely the selected weight, and thus the weighted vertices in the component. Notice that as the total weight in the component is k_ι, we have that $|Z| \leq k_\iota$.

For a set $A \subseteq [g]$ a *signature* is a tuple $(Z_{-\rho}, f_{-\rho}), \ldots, (Z_\eta, f_\eta)$, where $Z_{-\rho} \ldots Z_\eta$ form a non-trivial partition of A and for all $\iota \in [-\rho,\eta]$ and $f_\iota : Z_\iota \to \mathcal{W}_\iota$, we have $(Z_\iota, f_\iota) \in \mathcal{F}(A,\iota)$. Denote by $\mathcal{S}(A)$ the set of signatures of A. Notice A may be empty. We say that $A \subseteq [g]$ is *proper with respect to* G if $\mathcal{S}(A) \neq \emptyset$ and, for any $i \in [g] - A$, there is $\iota \in [-\rho,\eta]$ such that $\alpha(i,\iota,\mathbf{0}) = 1$.

The next step is to establish an equivalence relation to reduce the number of candidates in $[g]$ to form a proper set:

$$i \sim j \text{ iff } \ \forall \iota \in [-\rho,\eta] \, \forall W \in \mathcal{W}_\iota \ \ \alpha(i,\iota,W) = \alpha(j,\iota,W). \qquad (1)$$

Let $\mathcal{P}$ be the partition of $[g]$ defined by the above equivalence relation. Notice that we have at most 2^{k^*} equivalence classes and that each class can be represented by a binary string of length k^*. Consider the set $D \subseteq [g]$ obtained by keeping $k+1$ representatives from each class with more than k members, otherwise take all the elements in the class. Notice that D corresponds to a set of components of G whose cardinality *does not depend* on n. Set $\widehat{G} = \bigcup_{i \in D} G_i$.

Lemma 5.1. *Let (H,C,K) be a simple partially weighted graph. Given a graph G together with a (H,G)-list L. Then, G has a list (H,C,K)-coloring if and only if there is some $A \subseteq D$ that is proper with respect to $\widehat{G}$.*

Proof. Let χ be a list (H,C,K)-coloring of (G,L). For any $\iota \in [-\rho,\eta]$, let $B_\iota = \{i \mid \chi(V(G_i)) \subseteq V(H_\iota)\}$ be the set of indices of connected components of G that are mapped to $V(H_\iota)$. For any $i \in B_\iota$, let W_i be the partially weighted graph defined as $W_i(a) = |\chi^{-1}(a) \cap V(G_i)|$. Let $Z_\iota = \{i \in B_\iota \mid W_i \neq \mathbf{0}\}$. Clearly, $Z = \cup_{\iota \in [-\rho,\eta]} Z_\iota$ is proper.

For the reverse implication, if there is some $A \subseteq [g]$ that is proper, the partition of the connected components of G considered in one of the signatures in $\mathcal{S}(A)$ guarantees the existence of a valid list coloring that fills completely the weighted vertices. Note that the remaining components can be mapped to some component of H using zero weight, while respecting the list restrictions. So, we have that (G, L) has a list (H, C, K)-coloring. □

It follows from Lemma 5.1, that in order to check whether G has a list (H, C, K)-coloring, it suffices to compute the set D and check whether it contains a subset A that is proper with respect to $\widehat{G}$. Notice that this last check depends only on k and h and therefore it requires $O(f_2(k, h))$ steps. In order to compute D, we have to compute first the function α. If α is given, we can build the partition $\mathcal{P}$ of $[g]$ by applying g times the check in (1), which depends only on k and h. Therefore, the construction of D and $\widehat{G}$ can be done in time $O(g \cdot f_1^0(k, h))$, provided that the function α is known. To compute α, we first calculate the closed set $R_{i,\iota}$ and the kernel $\widehat{G}_{i,\iota}$ for the problem of asking whether there exists an (H_ι, C_ι, W)-coloring of G_i. This construction depends on the type of (H_ι, C_ι, W) and L (we use Lemmata 4.3, 4.6, and 4.9 respectively). Recall that only the type of (H_ι, C_ι, W) and L differentiates the way $R_{i,\iota}$ is being defined. Also, from the proof of Lemma 4.1, the choice of W is not important for the construction of these kernels as long as in the construction of the graph associated to $(G_i, R_{i,\iota})$ we keep $k + 1$ vertices for each "big" equivalence class P_v and $k + 1 \geq 1 + \sum_{u \in C_\iota} W(u)$ for any $W \in \mathcal{W}_\iota$. By Lemmata 4.3, 4.6, and 4.9, the construction of each $\widehat{G}_{i,\iota}$ requires $O((k + h)n_i + kh2^{h+k})$ steps where $n_i = |V(G_i)|$. Therefore, the computation of all kernels take at most $O((k+h)n + g \cdot kh2^{k+h}) = O((k+h)n + g \cdot f_1^1(k, h))$ steps. Going back to the computation of α, as long as we have the kernels, to compute $\alpha(i, \iota, W)$ for any $i \in [g], \iota \in [\gamma]$, and $W \in \mathcal{W}_\iota$ depends now only on k and h. However, as $1 \leq i \leq g$, we need $O(g \cdot f_1^2(k, h))$ steps to compute α. Choosing $f_1 \leq f_1^0, f_1^1, f_1^2$ we obtain the following result.

Theorem 5.2. *For simple partially weighted graph(H,C,K), the list (H,C,K)-coloring problem can be solved in time $O((k + h)n + f_1(k, h)g + f_2(k, h))$.*

6 The (H, C, K)-coloring Problem for Plain Partially Weighted Graphs

In this section we show the fixed parameter tractability of the (H, C, K)-coloring problem, when (H, C, K) is plain. In order to prove the result, we reduce the problem to an equivalent (H', C', K')-coloring problem in which (H', C', K') is simple.

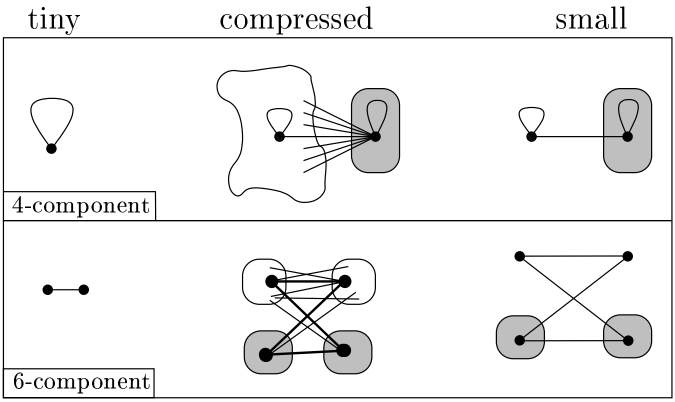


Fig. 5. Some particular cases of components in a plain (H, C, K)

We call a 4-component H' of H *compressed* if $|V(H') \cap C| = 1$. A 6-component is *compressed* if $|X(H') \cap C|, |Y(H') \cap C| \leq 1$. We say that a 4-component is *small* if it is compressed and $|V(H') - C| = 1$. A 6-component is *small* if it is compressed and $|V(H') - C| = 2$. Finally, we call a 4-component or a 6-component of H *tiny* if it is small and $C = \emptyset$. Figure 5 presents the different component subtypes.

A plain partially weighted graph (H, C, K) is said to be

- *compressed* if all the 4, and 6-components of (H, C, K) are compressed.
- *small* if all the 4 and 6-components of (H, C, K) are small.

Observe that if (H, C, K) is positive all its components are also positive. Furthermore, if (H, C, K) is plain and small then (H, C, K) is also simple.

The goal of this section is achieved through a series of parameterized reductions, from the general form of components to the *small* or *tiny* forms (see Figure 5). We assume that all the components of (H, C, K) have the form described in Section 2, even if the parameterized vertices have zero weight. Recall that $c = |C|$ and $h = |H|$.

Our first result allows us to consider only compressed plain partially weighted graphs.

Lemma 6.1. *For any plain partially weighted graph (H, C, K), there exists an equivalent (H', C', K'), which is compressed and plain. Moreover, such an assignment can be computed in $O(ch)$ steps.*

Proof. To prove the claim, we first show that for any plain partially weighted graph (H, C, K), where H contains $m \geq 1$ non-compressed 4-components, there exists an equivalent plain (H', C', K') such that H contains $m - 1$ non-compressed 4-components.

Let F be a non-compressed 4-component of H and set $D = C \cap V(F)$. Let $a_{\text{new}} \notin V(F)$ be a new vertex, and let, $k_D = \sum_{d \in D} K(d)$. We define

(F', C', K') in the following way:

$$\begin{aligned}
&V(F') = V(F) - D \cup \{a_{\text{new}}\},\\
&E(F') = E(F[V(F) - D]) \cup \{\{a, a_{\text{new}}\} \mid a \in N_H(D) - D\}\\
&\qquad\qquad\qquad\qquad \cup \{\{a_{\text{new}}, a_{\text{new}}\}\},\\
&C' = (C \cap V(F)) - D \cup \{a_{\text{new}}\}, \text{ and}\\
&K' \text{ is such that } K'|_{C-D} = K \text{ and } K'(a_{\text{new}}) = k_D.
\end{aligned}$$

Notice that replacing F by F' gives a plain partially weighted graph (H', C', K') that can be constructed in $O(ch|V(H')|)$ steps. Let us show that $(H', C', K') \sim (H, C, K)$. Given a graph G, if χ is a (H, C, K)-coloring of G, we define $\chi' : V(G) \to V(H')$ as

$$\chi'(v) = \begin{cases} \chi(v) \text{ if } \chi(v) \notin D,\\ a_{\text{new}} \text{ otherwise.}\end{cases}$$

then χ' is a (H', C', K')-coloring of G.

On the contrary, given an (H', C', K')-coloring χ' of G, then $|{\chi'}^{-1}(a_{\text{new}})| = k_D$. Therefore, we can define a function $\sigma : {\chi'}^{-1}(a_{\text{new}}) \to D$ such that, for all $d \in D$, we have $|\sigma^{-1}(d)| = K(d)$. Define $\chi : V(G) \to V(H)$ as:

$$\chi(v) = \begin{cases} \chi'(v) \text{ if } \chi'(v) \neq a_{\text{new}},\\ \sigma(v) \;\; \text{otherwise.}\end{cases}$$

Notice that χ is an (H, C, K)-coloring of G.

We now show that for any plain (H, C, K), where H contains $m \geq 1$ non-compressed 6-components, there exists an equivalent plain (H', C', K') such that H' contains $m - 1$ non-compressed 6-components.

Let F be a non-compressed 6-component of H and set $D_x = C \cap X(F)$ and $D_y = C \cap Y(F)$. Let $x_{\text{new}}, y_{\text{new}} \notin V(H)$ be two new vertices, and let, $k_x = \sum_{d \in D_x} K(d)$ and $k_y = \sum_{d \in D_y} K(d)$. We define (F', C', K') in the following way:

$$\begin{aligned}
&X(F') = X(F), Y(F') = Y(F) - D \cup \{x_{\text{new}}, y_{\text{new}}\},\\
&E(F') = E(F[V(F) - D_x - D_y]) \cup \{\{a, y_{\text{new}}\} \mid a \in N_H(D_x)\}\\
&\qquad\qquad \cup \{\{b, x_{\text{new}}\} \mid b \in N_H(D_y)\} \cup \{\{x_{\text{new}}, y_{\text{new}}\}\}\\
&C' = \{x_{\text{new}}, y_{\text{new}}\}, \text{ and}\\
&K' \text{ is such that } K'|_{C-D} = K, K'(x_{\text{new}}) = k_x. \text{ and } K'(y_{\text{new}}) = k_y.
\end{aligned}$$

Again replacing F by F' provides a plain partially weighted graph (H', C', K'), that can be constructed in $O(r|V(H')|)$ steps. Furthermore, it is easy to check that $(H', C', K') \sim (H, C, K)$.

The proof is complete, as we can iterate the construction until the point in which all the 4 and 6-components are compressed. ◻

The next result allows us to consider only small components.

Lemma 6.2. *For any compressed plain partially weighted graph* (H, C, K) *there exists an equivalent one,* (H', C, K)*, which is small and plain. Moreover, the assignment can be computed in* $O(h)$ *steps.*

Proof. Again we show that for any compressed plain partially weighted graph (H, C, K), with $m \geq 1$ non-small 4-components, there exists an equivalent compressed plain partially weighted graph (H', C, K) with $m - 1$ non-small 4-components. Later on, we show a similar result for the case of 6-components.

Suppose that F is a compressed 4-component of H such that $C \cap V(F) = \{d\}$ and d is adjacent in F to a looped non-weighted vertex x. Let H' be the graph obtained from H by replacing F with $F' = F[\{x, d\}]$. Then (H', C, K) is compressed and plain, has one non-small component less, and can be constructed in $O(V(H'))$ steps.

Let us show that the two partially weighted graphs are equivalent. Notice that any (H', C, K)-coloring of G is also a (H, C, K)-coloring of G, because F' is a subgraph of F containing all the parameterized vertices in F.

Let χ be an (H, C, K)-coloring of G, and consider the function $\rho : V(H) \to V(H')$ defined as follows:

$$\rho(a) = \begin{cases} a & \text{if } a \notin V(F), \\ d & \text{if } a = d, \\ x & \text{if } a \in V(F) - \{d\}. \end{cases}$$

Notice that ρ is an H'-coloring of H, and the function $\rho \circ \chi : V(G) \to V(H')$ is an (H', C, K)-coloring of G.

Suppose that F is a compressed 6-component of H such that for some $x, y \in V(F) - C$ we have that $C \cap V(F) = \{a, b\}$, a is adjacent in F to a vertex x, and b is adjacent in F to a vertex y. Furthermore, we also have that $\{x, y\} \in E(F)$. Let H' be the graph obtained from H by replacing F with $F' = (\{x, y, a, b\}, \{\{x, y\}, \{y, a\}, \{x, b\}, \{a, b\}\})$. Then (H', C, K) is compressed and plain, has one non-small 6-component less and can be constructed in $O(1)$ steps.

In a similar way to the case of 4-components, it can be shown (H', C, K) and (H, C, K) are equivalent.

As (H', C, K) contains less non-small 4 or 6-components, the lemma follows. ◻

Our next results establishes the equivalence with positive partial weight assignments.

Lemma 6.3. *For any plain small partially weighted graph* (H, C, K) *there exist an equivalent one* (H', C', K') *that is positive and small. Furthermore, it can be computed in* $O(ch)$ *steps.*

Proof. Consider the set $Z = \{c \in C \mid K(c) \neq 0\}$. Notice that $(H, C, K) \sim (H[Z], Z, K|_{C\cap Z})$, as it is not possible to map vertices in G to vertices outside Z. Furthermore, as we only remove parameterized vertices, $(H, C, K)[Z]$ is plain and small. □

Observe that a tiny component is $K_{1,0}$, $K_{1,1}$, or K_1^r. Note that $K_{1,0}$ can be mapped to $K_{1,1}$ and $K_{1,1}$ can be mapped to K_1^r. Therefore, it is enough to keep only one tiny component and preserve equivalence. We conclude with the following lemma.

Lemma 6.4. *For any plain, positive, and small partially weighted graph (H, C, K) there exists a plain, positive, and small partially weighted graph (H', C', K') equivalent to (H, C, K) that has at most one tiny component. Moreover, such an assignment can be computed in $O(h)$ steps.*

From Theorem 5.2 and by observing that a plain, positive and small, partially weighted graph is also simple, and that after applying the sequence of reductions we get a partially weighted graph in which each component has less than four vertices, we obtain the following result.

Theorem 6.5. *For any plain partially weighted graph (H, C, K), the(H, C, K)-coloring problem can be solved in $O(ch + (k+h)n + f_1(k, \gamma)g + f_2(k, \gamma))$ steps.*

7 Extension of the Results

7.1 Some Hardness Results

It is easy to show that there are partially weighted graphs in which the list (H, C, K)-coloring problem is NP-hard, taking H to be a non bi-arc graph and selecting $C = \emptyset$. In [DST01] it is also proved that the list (H, C, K)-coloring problem is NP-hard for any (H, C, K) in which $H - C$ is not a bi-arc graph, independently of the selection of C. In this section we produce a family of non-plain (H, C, K) for which the (H, C, K)-coloring problem is W[1]-hard. As we already mentioned in Section 1, this is considered a strong evidence that no fixed parameter algorithm exists for the list (H, C, K)-coloring problem for this family.

Assume that a partially weighted graph (H, C, K) has an un-looped vertex a in C and a looped vertex b in S, and that furthermore $\{a, b\} \in E(H)$. Given a graph G, by fixing the list of all the vertices in G to $\{a, b\}$, the corresponding list (H, C, K)-colorings correspond to the independent sets of G with k vertices, so we obtain the following result.

Theorem 7.1. *The list (H, C, K)-coloring problem is* W[1]*-hard if there is a looped vertex in $H - C$ connected to an un-looped vertex in C.*

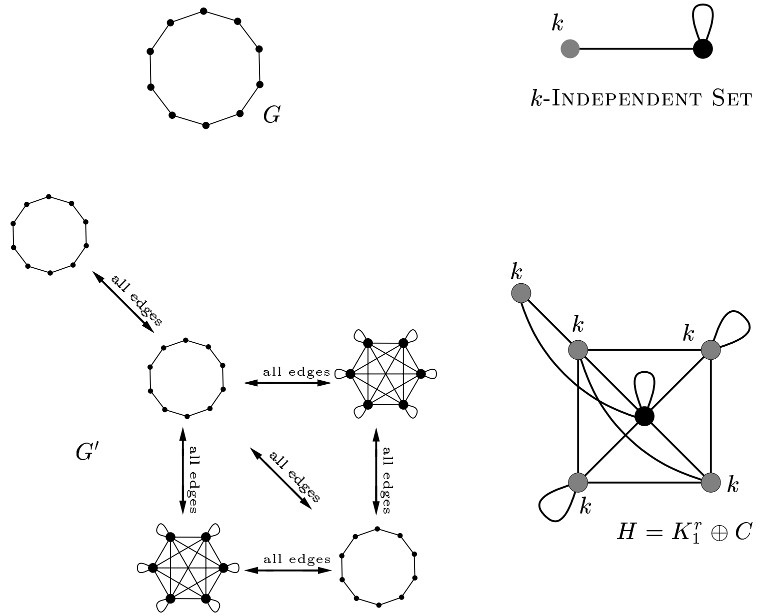


Fig. 6. The construction of the graph G' in the proof of Theorem 7.2

For the case of the (H, C, K)-coloring we can prove hardness for a particular case of graphs with an un-looped vertex. The difficulty in the reduction comes from the fact that now we cannot force any vertex to go in the desired position by setting the list.

Theorem 7.2. *The (H, C, K)-coloring problem is* W[1]*-hard, in the case that $H = K_1^r \oplus H'$ and $C = V(H')$, for some graph H' which contains at least one un-looped vertex.*

Proof. We provide a parameter preserving reduction from the k-independent set problem to the (H, C, K)-coloring, in the special case that $K = (k, \dots, k)$. Given a graph G we construct a new graph G' as follows: for every non-looped vertex v of F, set G_v to be a marked copy of G. For every looped vertex v of F, set G_v to be a marked copy of K_k^r. Furthermore, for every edge $\{u, v\} \in E(F)$ with $u \neq v$, add all the edges connecting vertices of G_u with vertices of G_v.

If G has an independent set of size k, then G' has trivially an (H, C, K)-coloring.

To prove the other direction, assume that there is an (H, C, K)-coloring φ of G'. Let $C_1 \subseteq C$ be the set of non-looped vertices in F. Observe that $\varphi^{-1}(C_1)$ contains only vertices from copies of G. Let D be a maximum independent set in $H[C_1]$, let $d = |D|$ and note that $d > 1$ as, by hypothesis, C_1 contains at least one vertex. We know that $|\varphi^{-1}(D)| = kd$ and that $\varphi^{-1}(D)$ is an independent set of G'. Let $I = \{v \in V(G) |$ there is a copy of v in $\varphi^{-1}(S)\}$, and for $v \in I$ let d_v be the number of copies of v in $\varphi^{-1}(D)$. By construction, if we take a set of vertices from more that d different copies of G, the set is not independent in G', because d is the size of the maximum independent set of C_1, and thus at least two of the copies are connected. Therefore, $d_v \leq s$. We know that $\sum_{v\in I} d_v = kd$ and that $d_v \leq d$, which implies that $|I| \geq k$. Therefore, G has an independent set of size at least k. ◻

7.2 Computing a (List) (H, C, K)-coloring

We describe now the changes to be performed in the decision algorithms described in Sections 3, 4, and 5, in order to compute a solution to the list (H, C, K)-coloring problem, if one exists. We use the empty function, $\varnothing$, to return the non-existence of a list (H, C, K)-coloring.

Algorithm Find-basic is obtained from algorithm Basic (Page 380) making it return $\varnothing$ when there is no list (H, C, K)-coloring, otherwise the coloring is obtained by selecting, for each vertex $v \in V(G)$, exactly one vertex in $L(v)$.

Algorithm Find-list-basic is obtained from algorithm List-basic (Page 382) making it return $\varnothing$ when appropriate. In the case that there is a list (H, C, K)-coloring, one of the list (H, C, K) colorings is defined by the pairs associated by the computed matching, together with a coloring on the unmatched vertices computed with an adequate call to Find-basic.

Algorithm Find-self-red is obtained from algorithm Self-red (Page 383) replacing the base calls to calls to the constructive algorithm. After a recursive call finishes, the algorithm checks whether it has provided a valid coloring. In such a case it adds to the computed coloring the last selected pair and returns it. In the case that all the recursive calls are unsuccessful, it returns $\varnothing$.

Algorithm Find-bipartite is obtained from algorithm Bipartite (Page 384) replacing the base calls to corresponding calls to the constructive algorithm for the other basic cases.

Using the results in Section 3 and the previous algorithms we have the following result:

Lemma 7.3. *Let (H, C, K) be a partially weighted graph. Given a graph G and an (H, G)-list L such that there is a a list (H, C, K)-coloring of (G, L), then*

- *When $E(G) = \emptyset$ or $H = K_h^r$, Algorithm Find-list-basic correctly computes in $O((k + h)n + nk \log k\sqrt{n + k})$ steps, a list (H, C, K)-coloring of (G, L).*
- *When $E(H - C) = \emptyset$, Algorithm Find-self-red correctly computes in $O(2^k c^k[(k + h)n + nk\sqrt{n + k} \log k]))$ steps, a list (H, C, K)-coloring of (G, L).*
- *When $H = K_{x,y}$, Algorithm Find-bipartite correctly computes in $O((k + h)n + nk\sqrt{n + k} \log k)$ steps, a list (H, C, K)-coloring of (G, L).*

For the case in which G is connected and (H, C, K) is a 1,2 or 3-component, we run the previous algorithms on the corresponding kernel $\widehat{G}$. As a corollary of Lemma 7.3, we can derive an algorithm Extend that given a list (H, C, K)-coloring of $(\widehat{G}, L)$ it obtains a list (H, C, K)-coloring of (G, L), in $O(n)$ steps. Therefore, by modifying the algorithms List-coloring, for each of the three types of components, changing the call to the basic algorithm for the call to the corresponding find algorithm and computing the extension of the obtained coloring, we have algorithms for computing a list (H, C, K) coloring if there is one. The time bounds are the same as for the decision problem.

Theorem 7.4. *Let (H, C, K) be a simple partially weighted graph. Given a connected graph G and an (H, G)-list L such that there is a a list (H, C, K)-coloring of (G, L). There are efficient fixed parameter algorithms for computing a list (H, C, K)-coloring of (G, L).*

The following procedure is a description of an algorithm that constructs a list (H, C, K)-coloring for the general case and uses the definitions and notation stated in Section 6 . We first enhance the function α in such a way that $\alpha(i, \iota, W)$ returns a list (H_ι, C_ι, W)-coloring of $(\widehat{G}_i, L_i)$, and it returns $\varnothing$ if such a coloring does not exist. We also enhance accordingly the definitions of $\mathcal{F}(A, \iota)$ and signature.

$$\mathcal{F}(A,\iota) = \{ \ (Z,f) \mid Z \subseteq A, f : Z \to \mathcal{W}_\iota^+, \\ \sum_{i \in Z} f(i) = K_\iota \wedge \forall i \in Z \ \alpha(i,\iota,f(i)) \neq \varnothing)\}.$$

The signature of a set $A \subseteq [g]$ is the tuple $(Z_{-\rho}, f_{-\rho}, \chi_{-\rho}), \ldots, (Z_\eta, f_\eta)$ where $Z_{-\rho} \ldots Z_\eta$ form a non trivial partition of A, and for all $\iota \in [-\rho, \eta]$ and $f_\iota : Z_\iota \to \mathcal{W}_\iota$ we have $(Z_\iota, f_\iota) \in \mathcal{F}(A, \iota)$. We compute the partition $\mathcal{P}$ and the set D as we did in Section 5 for the decision version, and check whether there exists some subset $A \subseteq D$ that is proper with respect to the graph $\widehat{G}$ formed by the union of the components with indices in D. If no such a subset exists, we return $\varnothing$. Otherwise, we compute one of the signatures of A, let us say $(Z_{-\rho}, f_{-\rho}), \ldots, (Z_\eta, f_\eta)$. Notice that, given the enhanced function α, we can also obtain a valid list in each block, furthermore this computation takes time that depends only on k and h. Then $\chi_A = \bigcup_{i \in A} \bigcup_{\iota \in [-\rho, \eta]} \chi_{i\iota}$ is a (H, C, K)-coloring of the graph formed by the union of the components in A. As A is proper, then for any $i \in D - A$ there is a $\iota_i \in [\rho, \eta]$ such that $\alpha(i, \iota_i, \mathbf{0}) = \chi_i$. If $\chi_{A^*} = \bigcup_{i \in D - A} \chi_i$, then $\chi_A \cup \chi_{A^*}$ is an (H, C, K)-coloring of $\widehat{G}$. We have to extend this coloring to a coloring of G. By construction of M, for each $j \in [g] - D$ there are at least $k + 1$ indices $i_1, \ldots, i_{k+1} \in D - A$, where $i_h \sim j$ for $h = 1, \ldots, k+1$. As we have chosen in D, $k+1$ representatives of the same class, one of them must be mapped to some component of H with zero weight. Therefore, for any $i \in [g] - A$ we can find a $\iota_i \in [\rho, \eta]$ such that $\alpha(i, \iota_i, \mathbf{0}) = \chi_i'$, which takes time that is linear on g and some function depending only on k and h. We set $\chi_{\bar{A}} = \cup_{i \in [g] - A} \chi_i'$ and return $\chi = \chi_A \cup \chi_{A^*} \cup \chi_{\bar{A}}$ as an (H, C, K)-coloring of G.

Theorem 7.5. *Let (H,C,K) be a partially weighted graph, given a graph G and an (H,G)-list L. The algorithm described in the previous paragraph computes a list (H,C,K)-coloring of (G,L), if there is one, in time $O((k+h)n + f_1(k,h)g + f_2(k,h)$, where f_1 and f_2 are the functions in Theorem 5.2.*

For the (H,C,K)-coloring problem, our approach performs a parameter preserving reduction that starts with a plain partially weighted graph and ends up with a simple partially weighted graph. Observe, that if we look at the coloring produced by the algorithm for simple partially weighted graphs, the pre-images of the non-weighted vertices should be maintained, while the pre-images of the weighted vertices must be redistributed among the original vertices. However, as the connection with the non-weighted vertices consists of all edges, we can separate them in whatever order to fill the weighted vertices. This leads us to the following result.

Theorem 7.6. *Let (H,C,K) be a plain partially weighted graph. Given a graph G that is (H,C,K)-colorable, an (H,C,K)-coloring of G can be computed in $O(ch + (k+h)n + f_1(k,\gamma)g + f_2(k,\gamma))$ steps, where f_1 and f_2 are the functions described in Theorem 6.5.*

7.3 The List $(H, C, \leq K)$-coloring

If in the definition of (H,C,K)-coloring we replace "=" with "≤", we get what we call a $(H,C,\leq K)$-*coloring* of G (see for ex. [DST01]). We can define again all the parameterized versions in the same way, replacing the notion of (H,C,K)-coloring for that of $(H,C,\leq K)$-coloring. Thus, the $(H,C,\leq K)$-*coloring* problem checks whether G has an $(H,C,\leq K)$-coloring. The list $(H,C,\leq K)$-coloring problem checks whether G has a list $(H,C,\leq K)$-coloring.

In a similar way, we can define the ≤-*equivalent* denoted by $(H,C,K) \sim_{\leq} (H',C',K')$. The following result is easy to prove,

Lemma 7.7. *Let (H,C,K) be a partially weighted graph, G a graph, L an (H,G)-list, and let $B = B(H,C,K,G,L)$ be the bipartite graph associated to (G,L). Assume that either G is edge-less, or $H = K_h^r$. Let $p = |\{v \in V(G) \mid L(v) \cap S = \emptyset\}|$. If $p > k$, there is no list $(H,C,\leq K)$-coloring of (G,L). Otherwise, there is a list $(H,C,\leq K)$-coloring of (G,L) if and only if the weight of a maximum matching in B is at most kp.*

Thus, Algorithm List-basic in Section 3, can be easily adapted to decide the existence of an $(H,C,\leq K)$-coloring, by changing the condition $w(M) < kp + k - p$ in the second **if** by the condition $w(M) < kp$. This change insures that the matching covers the set of vertices that must be mapped to C, although this assignment might not fill completely the vertices in C. We refer to such a modification as the Find-list-basic-leq function. Observe that the Basic algorithm, in Section 3, also works for the considered cases of the $(H, C, \leq K)$ coloring. The remaining basic algorithms perform calls either to the Basic or to the List-basic functions. We need to add the suffix leq for denoting the variation of the algorithms given in Section 5 in which the calls to List-basic are replaced by call to List-basic-leq. We have:

Lemma 7.8. *Let (H, C, K) be a simple partially weighted graph and let G be a connected graph and L an (H, G)-list.*

- *If G has no edges or $H = K_h^r$, List-basic-leq correctly checks whether there is a list $(H, C, \leq K)$-coloring of (G, L), in $O((k+h)n + nk\sqrt{n+k}\log k)$ steps,*
- *If $H - C$ is edge-less, Self-red-leq(H, C, K, G, L) correctly computes a list $(H, C, \leq K)$-coloring of (G, L), if there is one, and it runs in $O(2^k c^k[(k+h)n + nk\sqrt{n+k}\log k])$ steps,*
- *If $H = K_{x,y}$, Bipartite-leq(H,C,K,G,L) correctly computes a list $(H,C,\leq K)$-coloring of (G, L), if there is one, and it runs in $O((k+h)n + nk\sqrt{n+k}\log k)$ steps.*

For the general case we have to adapt the definition of the set D. After doing so, the rest of the algorithm is similar. In fact, we only have to modify the definition of $\mathcal{F}$ given in Section 5. Given $A \subseteq [g]$ and $\iota \in [-\rho, \eta]$, define

$$\mathcal{F}(A, \iota) = \{ \ (Z, f) \mid Z \subseteq A, f : Z \to \mathcal{W}_\iota^+, \sum_{i \in Z} f(i) \leq K_\iota \wedge \forall i \in Z \ \alpha(i, \iota, f(i)) = 1\}$$

The signature of a subset of $[g]$ and the partition $\mathcal{P}^{\leq}$ of $[g]$ are defined exactly as in Section 5. The same holds also for the construction of the set $D^{\leq}$ and the graph $\widehat{G}$. The following version of Theorem 5.1 holds.

Lemma 7.9. *Let (H, C, K) be a simple partially weighted graph. Given a graph G together with an (H, G)-list L, G has a list $(H, C, \leq K)$-coloring if and only if there is an $A \subseteq D^{\leq}$ that is proper with respect to $\widehat{G}$.*

Finally, by generating all the sets $A \subseteq D^{\leq}$ and all the elements in $\mathcal{F}(A)$ and performing the adequate test, we have,

Theorem 7.10. *Let (H, C, K) be a simple partially weighted graph. Then, the list $(H, C, \leq K)$-coloring problem can be solved in $O((k+h)n + f_1'(k, h)g + f_2'(k, h))$ steps.*

We have just proved that when (H, C, K) is simple, the list $(H, C, \leq K)$-coloring problem is in FPT. Observe that the reductions presented in Section 6 are also valid when considering the $(H, C, \leq K)$-coloring. Therefore, we conclude that the (H, C, K)-coloring problem is in FPT when (H, C, K) is plain.

Theorem 7.11. *Let (H, C, K) be a plain partially weighted graph. Then the $(H, C, \leq K)$-coloring problem can be solved in $O(ch + (k+h)n + f_1'(k, \gamma)g + f_2'(k, \gamma))$ steps.*

8 Conclusions and Open Problems

We have shown that for the family of simple partially weighted graphs, the list (H, C, K)-coloring problem has an efficient linear fixed parameter algorithm. We also have shown that for the family of plain partially weighted graphs, the (H, C, K)-coloring problem has an efficient linear fixed parameter algorithm.

As a future line of research it is worth to mention that the class of simple partially weighted graphs has been obtained by starting from a complete reflexive graph or a complete bipartite irreflexive graph and fixing the parameterized vertices. Next step is to look for other edges in the parameterized part that could be removed. It is clear from the hardness results that loops play a especial role, as by removing them from some of the weighted vertices of a reflexive graph or by adding them to some unweighted vertices of an irreflexive graph we get partially weighted graphs for which the list (H, C, K)-coloring is W[1]-hard. It remains to explore other classes of variations in which we add or remove edges.

Finally let us mention that recently the k-Almost Bipartite Graph (mentioned in the introduction) has been shown to belong to FPT [RSV04]. This problem can be reformulated as a (H, C, K)-coloring problem where $H = K_{1,1} \oplus K_1^r$ and $C = V(K_1^r)$ (see Figure 1). A natural extension is to consider $H = K_{x,y} \oplus K_c^r$ and put weight only in the reflexive clique. We conjecture that for such a partially weighted graph the (H, C, K)-coloring problems is also fixed parameter tractable.

Acknowledgement. We thank an anonymous referee for the careful comments which allowed us to improve the presentation and the writing quality and clarity.

References

[BG93] J.F. Buss and J. Goldsmith, Nondeterminism within P, *SIAM Journal on Computing*, 22(3):560–572, 1993.

[DST01] J. Díaz, M. Serna, and D. Thilikos, (H, C, K)-coloring: Fast, easy and hard cases, In J. Sgall, A. Pultr, and P. Kolman, editors, *Mathematical Foundations of Computer Science*, volume 2136, pages 304–315. Lecture Notes in Computer Science, Springer-Verlag, 2001.

[DST04a] J. Díaz, M. Serna, and D. M. Thilikos, Recent results on parameterized H-colorings, In J. Nešetřil and P. Winkler, editors, *Graphs, Morphisms and Statistical Physics*, DIMACS series in Discrete Mathematics and Theoretical Computer Science, vol 63, pp. 65–85. American Mathematical Society, 2004.

[DST05] J. Díaz, M. Serna, and D. Thilikos, The restrictive H-coloring problem, *Discrete Applied Mathematics*, 145:297–305, 2005.

[DST04b] J. Díaz, M. Serna, and D. Thilikos, Fixed parameter algorithms for counting and deciding bounded restrictive H-coloring, In S. Albers and

T. Radzic, editors, *Algorithms ESA 2004*, volume 3221, pages 275–286, Lecture Notes in Computer Science, Springer-Verlag, 2004.

[DST06$^+$] J. Díaz, M. Serna, and D. M. Thilikos, Complexity issues on Bounded Restrictive H-colorings, *Discrete Applied Mathematics*, to appear, 2005.

[DF99] R. G. Downey and M. R. Fellows, *Parameterized complexity*, Springer-Verlag, New York, 1999.

[DF95a] R. G. Downey and M. R. Fellows, Fixed-parameter tractability and completeness, I, Basic results, *SIAM Journal on Computing*, 24(4):873–921, 1995.

[DF95b] R. G. Downey and M. R. Fellows, Fixed-parameter tractability and completeness II: On completeness for $W[1]$, *Theoretical Computer Science*, 141(1-2):109–131, 1995.

[FH98] T. Feder and P. Hell, List homomorphisms to reflexive graphs, *Journal of Combinatorial Theory (series B)*, 72(2):236–250, 1998.

[FHH99] T. Feder, P. Hell, and J. Huang, List homomorphisms and circular arc graphs, *Combinatorica*, 19:487–505, 1999.

[FHH03] T. Feder, P. Hell, and J. Huang, Bi-arc graphs and the complexity of list homomorphisms, *Journal of Graph Theory*, 42:61–80, 2003.

[GT89] H. Gabow and R. Tarjan, Faster scaling algorithms for network problems, *SIAM Journal on Computing*, 18:1013–1036, 1989.

[HN90] P. Hell and J. Nešetřil, On the complexity of H-coloring, *Journal of Combinatorial Theory (series B)*, 48:92–110, 1990.

[HN04] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford University Press, 2004.

[RSV04] B. Reed, K. Smith, and A. Vetta, Finding odd cycle transversals, *Operation Research Letters*, 32:299-301, 2004.

From Graph Colouring to Constraint Satisfaction: There and Back Again

Pavol Hell

School of Computing Science, Simon Fraser University, Burnaby, B.C., V5A 1S6, Canada
pavol@cs.sfu.ca

Summary. Graph colourings may be viewed as special constraint satisfaction problems. The class of k-colouring problems enjoys a well known dichotomy of complexity – these problems are polynomial time solvable when $k \leq 2$, and NP-complete when $k \geq 3$. For general constraint satisfaction problems such dichotomy was conjectured by Feder and Vardi, but has still not been proved in full generality. We discuss some results and techniques related to this Dichotomy Conjecture. We focus on the effects of a new concept of 'fullness', and how it affects the complexity of constraint satisfaction problems and their dichotomy. Full constraint satisfaction problems may then be specialized back to graph colourings, yielding an interesting novel class of problems in graph theory, related to the study of graph perfection.

AMS Subject Classification. 05C15, 05C17, 05C75, 05C85, 08A70, 68Q25.
Keywords. Colourings, homomorphisms, constraint satisfaction problems, retractions, extensions, dichotomy conjecture, polymorphisms, full homomorphisms, matrix partitions, trigraphs, perfect graphs, polynomial algorithms, NP-completeness.

1 From Colouring to Constraint Satisfaction

We introduce the topic of this survey by means of an example from examination scheduling [CKKLS04, LS86, SO74]. Each semester at a typical North American university, the courses taught need to be scheduled for examination: several courses can be examined in one examination period, but two courses that have common students must be scheduled at different times. The university aims to schedule all examinations in as few examination periods as possible.

The simplifaction described above leads to a classical application of graph colouring [CKKLS04, LS86, SO74]. We form the graph G in which the vertices are the courses taught, and in which two courses are adjacent just if the courses *conflict*, i.e., have students in common. A schedule with k exam periods corresponds exactly to a k-colouring of G.

In a somewhat refined model of the situation, we may try to schedule the exams in a way which avoids a student taking two exams that are too close together. Thus certain pairs of exam periods are *incompatible* – for instance by being too close together in terms of timing. Now we may introduce a second graph, H, in which the vertices are the proposed exam periods, and in which two periods are adjacent just if they are compatible. We seek an assignment of courses to exam periods in which conflicting courses are assigned compatible periods. In terms of the graphs G and H, this is exactly a *homomorphism* of G to H, i.e., a mapping $f : V(G) \to V(H)$ such that $uv \in E(G)$ implies $f(u)f(v) \in E(H)$.

A much more realistic model of the situation is possible if, instead of requiring the assignment of courses to exam periods to preserve one binary relation, we require it to preserve any (finite) number of relations of arbitrary (finite) arity. This allows us to model, in addition to constraints ensuring that conflicting courses are scheduled at compatible exam periods (modeled above by binary relations), also more complex constraints. For instance, we may want to limit the number of experimental physics exams on any given day. If no more than, say, three such exams are permitted, we can model the constraint by quaternary relations: one quaternary relation R on the set of courses $V(G)$ has $(u, v, w, z) \in R$ just if all u, v, w, z are all courses in experimental physics, and one corresponding quaternary relation S on the set of exam periods $V(H)$ has $(a, b, c, d) \in S$ just if the periods a, b, c, d do not all take place on the same day. Then any mapping $f : V(G) \to V(H)$ which preserves these quaternary relations, in the sense that $(u, v, w, z) \in R$ implies $(f(u), f(v), f(w), f(z)) \in S$, satisfies the requirement that no four experimental physics courses are examined on the same day. Of course, we may still want to preserve the binary relations ensuring that conflicting courses receive compatible exam periods, and, in general, we may want to satisfy many different constraints at the same time. Constraint satisfaction problems can be viewed as seeking a mapping that preserves any number of relations.

1.1 Relational Structures

A *relational structure* G consists of a set $V(G)$, whose elements we shall call *vertices* in order to underscore our graph motivation, and a finite number of relations $R_1, R_2, \ldots, R_p$, of arities $r_1, r_2, \ldots, r_p$ respectively. The vector $(r_1, r_2, \ldots, r_p)$ is called *the type* of G. Given two relational structures G (with vertex set $V(G)$ and relations $R_1, R_2, \ldots, R_p$) and H (with vertex set $V(H)$ and relations $S_1, S_2, \ldots, S_p$), of the same type (the arity of each R_i is the same as that of the corresponding S_i), a *homomorphism* of G to H is a mapping $f : V(G) \to V(H)$ which preserves all pairs of corresponding relations, i.e., such that $(v_1, v_2, \ldots, v_{r_i}) \in R_i$ implies $(f(v_1), f(v_2), \ldots, f(v_{r_i})) \in S_i$, for all $i = 1, 2, \ldots, p$.

It should now be clear that the context of homomorphism of relational structures allows us to model simultaneously both the (binary) constraint ensuring that conflicting courses are scheduled at compatible times, and the (quaternary) constraint ensuring that no more than three experimental physics exams are scheduled on the same day, as well as any number of additional constraints. Before leaving our motivational example, let us illustrate the case of unary constraints, which will play a pivotal role in our discussions. Consider the following constraint: suppose all chemistry exams have to take place in the first ten days of the examination. This is best modeled by *unary constraints* R_i and S_i, i.e., constraints with arity $r_i = 1$. Recall that an r-ary relation on V is simply a subset of the r-tuple cartesian product $V \times V \times \ldots \times V$; hence a unary constraint on V is simply a subset of V. To model our restriction on the timing of the chemistry exams, we introduce the unary relation R on the set $V(G)$ of courses, where R is the set of all chemistry courses, and the unary relation S on the set $V(H)$ of exam periods, where S is the set of all the periods that take place in the first ten days. Now preserving the relations R and S amounts to scheduling all chemistry exams ($v \in R$) during the first ten days ($f(v) \in S$). The set of all unary relations of a relational structure H will be denoted by $\mathcal{U}(H)$.

1.2 Constraint Satisfaction Problems

A *constraint satisfaction problem* is the problem of existence of a homomorphism of a relational structure G to a relational structure H of the same type. This model is useful for applications typified by the above example, and many others in artificial intelligence, in particular in scheduling, planning, data bases, machine vision, belief maintenance, temporal reasoning, type reconstruction, and many other areas [CKS01, Dec92, Kum92, LM92, Mes89, Mit84, Tsa93, Var00, WO89]. The model was pioneered by Montanari [Mon74], in a slightly different terminology, where each r_i-tuple of each R_i is viewed as a separate 'constraint'; this is still the more predominant terminology in AI applications [Tsa93]. The formulation given above was first described by Feder and Vardi [FV98]. It plays the main role in the theoretical analysis of constraint satisfaction problems, such as [BJ01, FV98, Jea98]. The study of homomorphisms between relational structures has a long tradition in the 'Prague School' of category theory [HN04b], and many theoretical insights, such as recorded for instance in the monograph [PT80], are applicable to the present context.

A typical aspect of our example scheduling problem is the fact that the structure G of courses to be examined changes every semester, while the structure H of exam periods tends to be much more stable. (For instance, which courses are taught, and which pairs of courses have students in common, varies, while the total number of exam periods, the number of periods per day, and so on, may remain constant for several consecutive semesters.) A similar situation arises in other applications where the structure H tends to be fixed, and motivates the following model.

Let H be a fixed relational structure. The *constraint satisfaction problem* CSP(H) asks whether or not an input relational structure G, of the same type as H, admits a homomorphism to H. We note that when H and H' are two relational structures of the same type, which admit a homomorphism of H to H' as well as a homomorphism of H' to H, then the two problems CSP(H) and CSP(H') are obviously equivalent. (A structure G admits a homomorphism to H if and only if it admits a homomorphism to H', since the composition of two homomorphisms is again a homomorphism.) A relational structure H with no proper substructure H' to which H admits a homomorphism is called a *core*. It is easy to check that every structure H has a unique (up to isomorphism) core substructure H' to which it admits a homomorphism [HN92, HN04b]. Thus we typically restrict our attention to problems CSP(H) where H is a core.

We note that the problem CSP(H), where H has only one relation, S_1, that is binary and symmetric, corresponds to the problem of existence of a graph homomorphism of an input graph G to the fixed graph H, and is known as the *graph H-colouring problem* [HN90, HN04b]. When, moreover, S_1 is the relation of *non-equality* we obtain the *k-colouring problem* (where $k = |V(H)|$); this structure H is just the usual complete graph K_k.

2 Dichotomy

Constraint satisfaction problems are useful to model many applications; there are journals entirely devoted to solution of such problems, and books have been written about them [CKS01, Tsa93]. Our focus is on the theoretical aspects of constraint satisfaction.

The k-colouring problem is polynomial time solvable when $k = 1, 2$ and is NP-complete otherwise [GJ79]. This well known fact illustrates the *dichotomy* of possible complexities of the class of k-colouring problems, as k varies. There is, in principle, no reason for each colouring problem to be polynomial time solvable (one of the easiest problems in NP) or NP-complete (one of the hardest problems in NP). Indeed, Ladner [Lad75] has shown that if P $\neq$ NP, there are in NP problems that are neither polynomial nor NP-complete – in fact there must be an infinite hierarchy of such (non-polynomially-equivalent) problems. Since these *intermediate difficulty (ID) problems* must exist in NP (unless NP=P), dichotomies are always somewhat surprizing, especially if they apply to a broad class of problems.

Dichotomy for graph H-colouring problems has been proved by Jarik Nešetřil and the author in [HN90]. In other words, we have shown that, for every graph H, the problem of deciding the existence of a homomorphism of an input graph G to H is polynomial time solvable, or is NP-complete. More specifically, we have shown that the graph H-colouring problem is polynomial time solvable when H is bipartite or has a loop, and is NP-complete otherwise. Since this survey is dedicated to the celebration of Jarik's sixtieth birthday, I will record briefly my recollection of this collaboration. (The paper [HN90] is the most frequently cited publication for each of us.) In 1985, Jarik came for a six-month visit to SFU, and we agreed to try to prove this result, which had been conjectured in one of David Johnson's NP-completeness columns [Joh82]. What we didn't expect was that this would take the entire six months; and in fact we only completed the work by email, several months later. It is of course easy to show that any concrete nonbipartite irreflexive graph H yields an NP-complete H-colouring problem, yet to show it for all of them proved a challenge. Our paper is quite long and involved, and only recently has a shorter proof been found [Bul06$^+$]. In hindsight, it appears our effort was hampered by the narrow focus on graphs; by looking at the problem in the general framework of constraint satisfaction, one can facilitate the arguments by using relational structures of higher arities in the intermediate steps. It also helps to use the algebraic techniques discussed below [BJ01, BJK05, Jea98].

2.1 The Dichotomy Conjecture

Dichotomy for the general class of problems CSP(H) is not known, and has been conjectured by Feder and Vardi [FV98].

Conjecture 2.1 (The Dichotomy Conjecture). *For any relational structure H, the problem CSP(H) is NP-complete or polynomial time solvable.*

The above dichotomy for graph H-colouring problems was only the secondary motivation for the Dichotomy Conjecture [FV98]. The primary motivation was actually a result of Schaeffer [Sch78], classifying all Boolean satisfiability problems as NP-complete or polynomial time solvable. These are the problems CSP(H) where H has two vertices, say, 0 and 1. Schaeffer's result establishes the Dichotomy Conjecture in this case. To describe Schaeffer's classification, we shall recall four well known operations on tuples. The *OR* operation on two tuples $(a_1, a_2, \ldots, a_s)$ and $(b_1, b_2, \ldots, b_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each $z_i = a_i \vee b_i$ ($z_i = 1$ unless both $a_i = b_i = 0$, in which case $z_i = 0$). The *AND* operation on two tuples $(a_1, a_2, \ldots, a_s)$ and $(b_1, b_2, \ldots, b_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each $z_i = a_i \wedge b_i$ ($z_i = 0$ unless both $a_i = b_i = 1$, in which case $z_i = 1$). The *MAJORITY* operation on three tuples $(a_1, a_2, \ldots, a_s)$, $(b_1, b_2, \ldots, b_s)$, and $(c_1, c_2, \ldots, c_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each z_i is the majority value (0 or 1) of a_i, b_i, c_i. The *XOR* (exclusive OR, also known as MINORITY) operation on three tuples $(a_1, a_2, \ldots, a_s)$, $(b_1, b_2, \ldots, b_s)$, and $(c_1, c_2, \ldots, c_s)$ is the tuple $(z_1, z_2, \ldots, z_s)$ where each z_i is the exclusive-or value of a_i, b_i, c_i (equal to 1 if the number of 1's amongst a_i, b_i, c_i is odd, and 0 otherwise). Schaeffer proved the following classification [Sch78].

Theorem 2.2. *Suppose H is a relational structure with $V(H) = \{0, 1\}$ and relations $S_1, S_2, \ldots, S_p$. Then CSP(H) is NP-complete, except in the following, polynomial time solvable, cases:*

1. *each* S_i *contains the* s_i*-tuple* $(0, 0, \ldots, 0)$*; or*
2. *each* S_i *contains the* s_i*-tuple* $(1, 1, \ldots, 1)$*; or*
3. *each* S_i *is closed under the OR operation; or*
4. *each* S_i *is closed under the AND operation; or*
5. *each* S_i *is closed under the MAJORITY operation; or*
6. *each* S_i *is closed under the XOR operation.*

The polynomial time algorithms for these cases are well known. In cases 1 and 2, the core of H has one vertex, and any structure G admits a homomorphism this core (and hence to H). Case 3 (respectively 4) corresponds to problems equivalent to the case where each S_i consists of all s_i-tuples with 1 in the first coordinate, plus possibly the tuple $(0, 0, \ldots, 0)$ (respectively all s_i-tuples with 0 in the first coordinate, plus possibly the tuple $(1, 1, \ldots, 1)$). Thus they can be expressed by *Horn clauses* (respectively *dual-Horn clauses*), i.e., disjunctions with at most one negated (respectively unnegated) variable, and solved as in [DG84, Horn51]. Case 5 corresponds to problems equivalent to H having just four binary relations, S_1 consisting of all pairs other than $(0, 0)$, S_2 consisting of all pairs other than $(0, 1)$, S_3, consisting of all pairs other than $(1, 0)$, and S_4, consisting of all pairs other than $(1, 1)$. Thus these are the problems expressible by disjunctions with two variables each, i.e., by 2-satisfiability [APT79]. The last case corresponds to systems of linear equations modulo two, solvable by Gaussian elimination.

In the intervening years, many other dichotomy results have been proved [BH90, BHM88, BHM92, Bul03, Bul02a, Dal00, DF03, Fed01b, Fed06$^+$, FF06$^+$, FH98, FHH99a, FHH99b, FHH06$^+$, FHH04, FHM03, FMS04, LZ03, ŁN03b, PT96], cf. [Fed04, HN04b], including an extension of Schaeffer's dichotomy theorem to all problems CSP(H) for structures H with up to three vertices [Bul02b].

The Dichotomy Conjecture appears to be an important and difficult question. In a sense, this class of problems is a largest class in which dichotomy can be expected. Specifically, Feder and Vardi [FV98] introduced two logic complexity classes A, B (the 'monotone strict NP without inequality' and the 'monadic strict NP without inequality'), each of which must contain intermediate difficulty (ID) problems (problems that are neither NP-complete nor polynomial time solvable, unless P=NP), yet such that each problem in their intersection (the 'monotone monadic strict NP without inequality') is polynomially equivalent to CSP(H) for some structure H, cf. Figure 1.

It is worth noting that dichotomy is *not* known for the case of *digraphs*, i.e., for problems CSP(H) where H has only one relation, S_1, which is binary (but not necessarily symmetric). Many results have been proved, classifying the complexity of these *digraph H-colouring problems* for special families of digraphs H [BH90, BHM95, BHM88, BHM92, Fed01b, FHM03, GWW92]. For instance, dichotomy is known to hold for the cases when the underlying undirected graph of H is complete [BHM88], or a path [GWW92], or a cycle [Fed01b]. In the case when H does not contain vertices of indegree or outde-

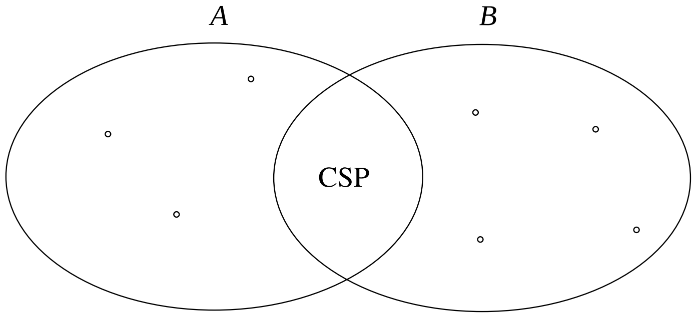


Fig. 1. Where the ID problems seem to lie

gree zero, a concrete graph theoretic classification has been conjectured for the complexity of digraph H-colouring problems [BH90]. (For an equivalent formulation, see [BHM95]; a special case of interest, where the underlying undirected graph of H is bipartite, is discussed in [BHM92].)

Conjecture 2.3. *Suppose H is a core digraph with all indegrees and outdegrees positive. If each component of H is a directed cycle the digraph H-colouring problem is polynomial time solvable; otherwise it is NP-complete.*

With vertices of indegree or outdegree zero, very little is known. For instance, dichotomy is not known if the underlying undirected graph of H is a tree, not even if it is a union of three paths meeting at one vertex [HNZ96].

Finally, we note that it was shown by Feder and Vardi [FV98] that dichotomy for digraph H-colouring problems would imply the entire Dichotomy Conjecture [FV98]. Thus there is a surprizing difference between the H-colouring problems for graphs and for digraphs.

2.2 Polymorphisms

The greatest progress on the dichotomy conjecture resulted from an algebraic approach pioneered by Jeavons [Jea98]. It turns out that what determines whether a structure H has a hard or easy constraint satisfaction problem CSP(H), is its set of polymorphisms. A *polymorphism (of order k)* of H is a mapping $f : V(H^k) \to V(H)$, such that $(v_1^j, v_2^j, \ldots, v_{r_i}^j) \in S_i$ for $j = 1, 2, \ldots, k$ implies that

$$(f(v_1^1, v_1^2, \ldots, v_1^k), f(v_2^1, v_2^2, \ldots, v_2^k), \ldots, f(v_{r_i}^1, v_{r_i}^2, \ldots, v_{r_i}^k)) \in S_i,$$

for all relations S_i of H. (Thus a polymorphism of order k of H is just a homomorphism of a suitably defined direct product power H^k to H [PT80].) Denoting by $\mathcal{P}ol(H)$ the set of all polymorphisms of H, Jeavons [Jea98] observed the following fact.

Theorem 2.4. *Assume $V(H) = V(H')$. If $\mathcal{P}ol(H') \subseteq \mathcal{P}ol(H)$, then CSP$(H)$ is polynomially reducible to CSP(H').*

In particular, if H and H' have exactly the same set of polymorphisms (of all orders), then the constraint satisfaction problems CSP(H) and CSP(H') are polynomially equivalent, regardless of the relations H and H' may have. The structures H and H' need not even have the same type. Consider, for example, Schaeffer's classification, Theorem 2.2, above. We have stated the result in a form which emphasizes that what makes the problems CSP(H) polynomial time solvable, is the existence of certain operations under which all the relations of H are closed. Each of these operations corresponds to a polymorphism of H, as we shall discuss below.

Thus structures H having many polymorphisms are likely to have polynomial time solvable problems CSP(H), and structures H which have few polymorphisms can be expected to have NP-complete problems CSP(H). Note that every structure H admits some polymorphisms – at least the projections π_i taking $(v_1, v_2, \ldots, v_k)$ to v_i. If the structure H has any automorphisms, i.e., bijective homomorphisms f of H to itself, then any composition $f \circ \pi_i$ is also a polymorphism of H. A structure H which has no other polymorphisms is called *projective*. It is known that if a structure H is projective, then the problem CSP(H) is NP-complete [Jea98, JCG97]. (It is worth noting that our definition of projectivity implies that the structure is a core, because we apply the requirement also to polymorphisms of order 1.)

One class of structures that are known to be projective are the complete graphs K_k, $k \geq 3$, i.e., the structures H with one binary relation, of non-equality cf., e.g., [LT02]. (Note that any permutation of the vertices of K_k is an automorphism of K_k.)

Theorem 2.5. *If $k \geq 3$, then K_k is projective.*

Corollary 2.6. *If $k \geq 3$, then the problem CSP(K_k) is NP-complete.*

The problem CSP(K_k) is of course just the well-known problem of k-colourability. Consider next the structure N which has two vertices $0, 1$ and one ternary relation $E(N) = \{(1,0,0), (0,1,0), (0,0,1), (1,1,0), (1,0,1), (0,1,1)\}$. The following folklore result, cf. [Sch78], makes a nice exercise [HN04b].

Theorem 2.7. *The structure N is projective.*

Corollary 2.8. *The problem CSP(N) is NP-complete.*

The problem CSP(N) takes as input a structure G with triples (u, v, w); its vertices must be assigned values $0, 1$ so that each triple receives at least one 0 and at least one 1. This is the problem of Not-All-Equal Three-Satisfiability Without Negated Variables, another standard NP-complete problem [GJ79].

From all available evidence one may guess that if a problem CSP(H) cannot be shown NP-complete by an (iterated) reduction from CSP(K_k) or CSP(N), then CSP(H) is polynomial time solvable. The concrete classification proposed in the next subsection [BJK05] is related to this intuition. On the other hand, it was shown by Luczak and Nešetřil [ŁN03b] that almost all structures are projective, and hence have NP-complete problems CSP(H). So, in this sense, the Dichotomy Conjecture is asymptotically true (and even asymptotically trivial).

2.3 A Proposed Dichotomy Classification

Let us now return to the idea that polymorphisms of H tend to make the problem CSP(H) polynomial time solvable. For the purposes of this discussion, we shall focus on polymorphisms f that are *idempotent*, i.e., satisfy $f(x, x, \ldots, x) = x$ for all vertices $x \in V(H)$. In fact, we may focus on structures H that only admit polymorphisms that are idempotent. We may, for instance [LZ04], assume that the unary relations $\mathcal{U}(H)$ include all constants $\{v\}, v \in V(H)$. (In particular, this implies that H is a core.) Such a structure H is projective if and only if it admits no polymorphism other than a projection. These (projective) structures H lead to NP-complete problems CSP(H), as discussed above. Admitting any (idempotent) polymorphism which is not a projection is sometimes sufficient to ensure that CSP(H) is polynomial time solvable. This is the case, for instance, when H has only two vertices, as can be checked from Schaeffer's classification, Theorem 2.2. However, it is not the case for general structures H [Bul06$^+$]. On the other hand, it turns out that admitting a polymorphism that is sufficiently far from being a projection, may help. We say that a polymorphism f is *inclusive in position* i, if it satisfies an identity involving two variables, with different entries in position i. More precisely, there exist choices $u_j, v_j \in \{u, v\}, j = 1, 2, \ldots, k$, with $u_i \neq v_i$, such that the identity

$$f(u_1, u_2, \ldots, u_k) = f(v_1, v_2, \ldots, v_k)$$

holds for all $u, v \in V(H)$. Polymorphisms inclusive in each position are called *Taylor operations* [LZ04]. Clearly, the i-th projection is not inclusive in position i, so projections are not Taylor operations, and, in fact, a Taylor operation is just what we sought – a polymorphism far enough from a projection. Indeed, it is still the case that if H does not admit a Taylor operation, the problem CSP(H) is NP-complete [BJK05, LZ04]. Furthermore, all known structures H that admit a Taylor operation turn out to have polynomial time solvable problem CSP(H). Thus the following conjecture has been proposed by Bulatov, Jeavons, and Krokhin [BJK05]; this particular formulation is from [LZ04]. Recall that we make the (technical) assumption that H only admits idempotent polymorphisms.

Conjecture 2.9. *If H admits a Taylor operation, then CSP(H) is polynomial time solvable. Otherwise CSP(H) is NP-complete.*

This is a refinement of the Dichotomy Conjecture of Feder and Vardi, since it postulates a classification (of sorts) of which problems CSP(H) are NP-complete and which are polynomial time solvable. For instance, the recent proof in [Bul06$^+$] of the dichotomy classification of graph H-colouring [HN90], actually implies the conjecture in that case (when H is a graph); in other words, all graphs that are not bipartite and have no loops are shown to not admit any Taylor operation.

As stated above, the NP-completeness part of the conjecture is known. It only remains to verify that the existence of a Taylor operation ensures polynomial time solvability. This is known for several classes of Taylor operations.

For instance, consider the following polymorphism that is often helpful. A *majority operation* is a polymorphism f of order three which has the *majority property*, that $f(u,u,v) = f(u,v,u) = f(v,u,u) = u$, for any vertices u and v. Clearly, a majority operation is inclusive in each position. Majority operations are fairly common. For instance, if H is a digraph (structure with one binary relation), whose underlying undirected graph is a path, then H admits a simple majority operation - namely $f(u,v,w)$ being the middle of the vertices u,v,w on the path. It is a simple exercise to verify that this definition (which clearly satisfies the majority property) yields a polymorphism of H; we simply check that in all cases when uu', vv', ww' are arcs of H, the middle vertices $f(u,v,w)f(u',v',w')$ also form an arc of H. For another example, consider the MAJORITY operation on Boolean tuples, defined just above Theorem 2.2. It is easy to check that MAJORITY defines a polymorphism (and hence a majority operation) on H if and only if each S_i is closed under MAJORITY. The following result goes back to (the STOC 1993 version of) [FV98]; cf. also [JCG97].

Theorem 2.10. *If H admits a majority operation, then CSP(H) is polynomial time solvable.*

The theorem generalizes to polymorphisms of higher orders [FV98], cf. [BFHHM04]. We apply the theorem to derive the following result. Here H is a digraph, i.e., a structure with one binary relation.

Corollary 2.11. *If the underlying undirected graph of H is a path, then the H-colouring problem is polynomial time solvable.*

This is a nontrivial result, first proved by a different technique in [GWW92]. It is, of course, trivial to decide whether or not a given digraph has a homomorphism to a fixed *directed* path, since once a first vertex v has been mapped (and there is only a fixed number of possible images for v), the images of all outneighbours and inneighbours of v are forced, and hence so are the images of all the vertices in the same component of G. However, when the underlying path of H has some arcs oriented one way and some another, the image of v may have two outneighbours and it is not clear to which of them should an outneighbour of v be mapped.

Before leaving this topic, we shall mention two other kinds of Taylor operations that are known to imply polynomial time algorithms. A *Maltsev operation* is a polymorphism f of order three which satisfies $f(u,u,v) = f(v,u,u) = v$ for all vertices u, v. A *semilattice operation* is a polymorphism f of order two which satisfies $f(u,u) = u, f(u,v) = f(v,u)$, and $f(a, f(b,c)) = f(f(a,b),c)$, for all vertices u, v. A semilattice operation is clearly inclusive in each position; to see that a Maltsev operation is inclusive in the second position consider the identity $f(u,u,v) = f(v,v,v)$. Recall our Boolean operations OR, AND, XOR, from Theorem 2.2. If each relation S_i in H is closed under OR (respectively AND, respectively XOR), these operations define a polymorphism of H, which is a semilattice operation for OR, AND and a Maltsev operation for XOR. It is known that the existence of either a Maltsev or a semilattice operation on a structure H ensures that CSP(H) is polynomial time solvable [Jea98, JCG95].

3 Full Constraint Satisfaction Problems

At this point in time, the classification conjecture of Bulatov, Jeavons, and Krokhin, and the Dichotomy Conjecture of Feder and Vardi, remain open, despite many partial results [BH90, BHM88, BHM92, Bul03, Bul02a, Bul02b, Dal00, DF03, Fed01b, Fed06$^+$, FF06$^+$, FH98, FHH99a, FHH99b, FHH06$^+$, FHM03, FMS04, HN90, LZ03, ŁN03b, PT96, Sch78], cf. [Fed04, HN04b].

We now introduce the notion of fullness [FH06$^+$a], which substantially changes the picture. Let Λ be a set of positive integers. A relational structure G is said to be Λ-*full* if for each integer $\ell \in \Lambda$ and every ℓ-tuple $(v_1, v_2, \ldots v_\ell)$ of vertices of G, there is a unique ℓ-ary relation R_i in G, and a unique permutation π such that $(v_{\pi(1)}, v_{\pi(1)}, \ldots, v_{\pi(1)}) \in R_i$. In other words, each ℓ-tuple of vertices is involved in a unique ℓ-ary relation, and in a unique way. (The uniqueness turns out not to be too important, but we include it for technical reasons.)

The restriction of the constraint satisfaction problem CSP(H) to Λ-full input structures G is denoted by $\text{CSP}_\Lambda(H)$. Note that this is *not* in general a constraint satisfaction problem (that is, it is not a problem CSP(H) for some structure H), and hence we have no a priori reason to expect dichotomy in this case. Indeed, it is possible that a restriction of an NP-complete problem becomes easier than NP-complete but not quite polynomial time solvable.

3.1 1-full Problems

When $\Lambda = \{\lambda\}$, we simply speak of λ-full structures G and the problems $\text{CSP}_\lambda(H)$. We first focus on the case when $\Lambda = \{1\}$. This restricts CSP(H) to 1-full input structures – that is input structures G in which each vertex is involved in a unique unary relation. If v is in the relation $R_i \in \mathcal{U}(G)$, then the corresponding relation $S_i \in \mathcal{U}(H)$ is the set of allowed images for v, under a homomorphism of G to H. These kinds of restrictions have been traditionally viewed as lists: each vertex v is given a *list* (set) $L(v)$ of allowed images, and a *list homomorphism* of G to H is a homomorphism f of G to H which satisfies $f(v) \in L(v)$ for all vertices v. As noted above, we let the list $L(v)$ be the set $S_i \in \mathcal{U}(H)$, where i is the subscript such that $v \in R_i$, $R_i \in \mathcal{U}(G)$.

Let us first consider the case when $\mathcal{U}(H)$ consists of all $2^{|V(H)|}$ subsets of $V(H)$. Such structures H are called *conservative*. This means that the lists on the vertices of G are allowed to be any sets of vertices of H. In this case, the problem $\mathrm{CSP}_1(H)$ is also called a *list constraint satisfaction problem*. It is easy to see that for conservative structres H, the problems $\mathrm{CSP}_1(H)$ and $\mathrm{CSP}(H)$ are in fact polynomially equivalent. Indeed, assume that S_k is the unary relation $V(H)$ on H: any vertex v of G which is in the relation R_k has the list $V(H)$ and hence is not restricted by a unary constraint. Thus every instance of $\mathrm{CSP}(H)$ can be transformed to a corresponding instance of $\mathrm{CSP}_1(H)$, in which all vertices are placed in the relation R_k. Since $\mathrm{CSP}_1(H)$ is also a restriction of $\mathrm{CSP}(H)$, we see that the two problems are polynomially equivalent. Therefore, the Dichotomy Conjecture does apply in this situation, and has in fact been verified by Bulatov [Bul03].

Theorem 3.1. *If H is a conservative structure, then $CSP_1(H)$ (and $CSP(H)$) is NP-complete or polynomial time solvable.*

This result has been motivated by a number of partial results proving dichotomy for classes of conservative structures [CCJ94, FH98, FHH99a, FHH99b]. In fact, in some cases we have nice classification results.

We shall consider structures H which consist of one binary relation and some (usually all) unary relations. We denote by H^- the digraph obtained by deleting all unary relations of H. Note that if the binary relation of H is symmetric, then H^- is a graph with loops allowed; if the binary relation of H is irreflexive, then H^- is a graph without loops; and if the binary relation of H is reflexive, then H^- is a graph in which each vertex has a loop.

The following classification is proved in [FHH99b]; bi-arc graphs are defined below.

Theorem 3.2. *Assume H is conservative, and the binary relation of H is symmetric (but not necessarily irreflexive). If H is a bi-arc graph then $CSP_1(H)$ (and $CSP(H)$) is polynomial-time solvable. Otherwise both problems are NP-complete.*

Let C be a fixed circle with two antipodal points p and q. A graph H (with loops allowed) is called a *bi-arc graph* if there exist on C arcs N_x, S_x, $x \in V(H)$, such that

- N_x contains p but not q and S_x contains q but not p,
- N_x intersects S_y if and only if N_y intersects S_x,

and such that for any $x, y \in V(H)$, not necessarily distinct,

- x and y are adjacent in H if and only if N_x is disjoint from S_y.

This class of graphs conveniently generalizes both the class of interval graphs, and the class of (complements of) circular arc graphs of clique covering number two [FHH99b, Gol80]. In particular, if the binary relation on H is reflexive (H^- is a graph with loops), the result simplifies as follows [FH98].

Corollary 3.3. *Assume H is conservative, and H^- a reflexive graph. If H is an interval graph then $CSP_1(H)$ (and $CSP(H)$) is polynomial-time solvable. Otherwise both problems are NP-complete.*

For irreflexive graphs, a similar simplification states that $\mathrm{CSP}_1(H)$ (and $\mathrm{CSP}(H)$) is polynomial-time solvable if H^- is bipartite and its complement $\overline{H^-}$ is a circular arc graph, and otherwise both problems are NP-complete [FHH99a].

3.2 Digraphs

The proof of Theorem 3.1 also involves a classification of sorts – the problem $\mathrm{CSP}_1(H)$ (and $\mathrm{CSP}(H)$) is NP-complete, unless H admits a polymorphism f such that, for any two vertices u, v of H, the restriction of f to tuples of u's and v's is equal to one of the following three types of polymorphisms: the majority operation, the semilattice operation, or the Maltsev operation (as discussed above). If such a polymorphism exists, then both problems are polynomial time solvable. Neverthless, it would still be worthwhile to obtain more combinatorial classifications of the sort illustrated in Theorem 3.2 and Corollary 3.3. In particular, it would be nice to have such a classification when H^- is a digraph. In [FHH04], we have conjectured such a classification when H^- is a reflexive digraph, and when H^- is an irreflexive digraph (cf. [FHK04]). (These can be viewed as *list digraph H-colouring problems.*)

Conjecture 3.4. *Let H be a conservative structure such that H^- is a reflexive digraph. If H^- admits an X-underbar enumeration, then $CSP_1(H)$ (and $CSP(H)$) is polynomial time solvable. Otherwise, both problems are NP-complete.*

An *X-underbar enumeration* of a digraph H is an ordering of its vertices as $v_1, v_2, \ldots, v_n$ such that whenever the arcs $v_i v_j$ and $v_{i'} v_{j'}$ are in H, then the arc $v_{i''} v_{j''}$, with $i'' = \min(i, i'), j'' = \min(j, j')$, is also in H. (The existence of an X-underbar enumeration can be replaced by the existence of a semilattice operation [FHH04].)

There is also a formulation similar to Corollary 3.3. Say that a reflexive digraph H is a *bi-interval digraph* if every vertex v can be assigned two real intervals I_v and J_v, with the same left endpoint l_v, such that the arcs of H are precisely the pairs uv such that either l_v is in I_u or l_u is in J_v. It is easy to see that reflexive digraph H has an X-underbar enumeration if and only if it is a bi-interval digraph. Thus our conjecture is that $\text{CSP}_1(H)$ (and $\text{CSP}(H)$) are polynomial time solvable for bi-interval digraphs and NP-complete otherwise [FHH04].

For irreflexive graphs, we have the following conjecture [FHH04].

Conjecture 3.5. *Let H be a conservative structure such that H^- is an irreflexive digraph. If H admits a majority operation, then $CSP_1(H)$ (and $CSP(H)$) is polynomial time solvable. Otherwise, both problems are NP-complete.*

In both cases, the polynomial time algorithms are well known. Several partial results supporting the NP-completeness statements of the conjectures can be found in [FHH04, FHK04, Lar05]. In particular, the conjectures have been verified for digraphs H with up to three vertices [FHK04], and for various families of digraphs, such as those digraphs whose underlying graphs are trees, complete graphs, or triangle-free graphs [FHH04, Lar05].

3.3 Non-conservative Structures and Retraction Problems

Recall that we denote by $\mathcal{U}(H)$ the set all unary relations of a structure H. Let us next consider the situation when $\mathcal{U}(H)$ does not contain all $2^{|V(H)|}$ subsets of $V(H)$. (Note that as long as $V(H) \in \mathcal{U}(H)$, it is still the case that $\text{CSP}_1(H)$ and $\text{CSP}(H)$ are polynomially equivalent.)

A classical problem of this type is the *retraction problem* (or the *precolouring extension problem*) [BDS87, BFH93, BP91, BHT92, Bor31, Fed04, FH03, Hell72, Hell74a, Hell74b, HR87, HT93, HT96, JMP86, Lar05, LZ04, Marx03, Pes88, Qui85, Vik03], which is the problem $\text{CSP}_1(H)$ with $\mathcal{U}(H)$ consisting of the set $V(H)$, and all singletons $\{x\}, x \in V(H)$. The effect of the singleton lists is to pre-define the colours of certain vertices, and then an extension of this 'pre-colouring' is sought. Alternately, if H is a substructure of G, and if the lists of each vertex $h \in V(H)$ is $\{h\}$ while all other lists are $V(H)$, then we seek a homomorphism f of G to its substructure H such that $f(h) = h$ for all $h \in V(H)$. Such homomorphisms are called *retractions* and they have been of interest since the time of Borsuk [Bor31, Hell72, JMP86, LZ04, Pes88, Qui85].

This problem has been studied mostly for structures H which contain, in addition to the unary relations $\mathcal{U}(H)$, only one binary relation, and thus define the digraph (often a partial order) or graph H^- as above. Many partial results on dichotomy of such retraction problems can be found in [FH03, BDS87, BFH93, BP91, FH03, FH98, FHH99a, Lar05, Pes88]; nevertheless, dichotomy is *not* known for retraction problems, and it was shown by Feder and Vardi [FV98] that dichotomy for retraction problems when H^- is a graph (or even a bipartite graph) would imply the entire Dichotomy Conjecture.

Many other natural choices of $\mathcal{U}(H)$ seem to suggest themselves for investigation. We close this section with a typical example from [FH98], highlighting the role of chordal graphs (for more on this topic, see [Fed04, Marx03]).

Theorem 3.6. *Let H^- be a reflexive graph and let $\mathcal{U}(H)$ consists of all sets $S \subseteq V(H)$ which induce a connected subgraph of H^-. If H^- is a chordal graph, then $CSP_1(H)$ is polynomial time solvable. Otherwise it is NP-complete.*

Note that if H is itself connected, the same result applies to CSP(H).

3.4 {1, 2}-full Problems

Consider next the case when $\Lambda = \{1, 2\}$. This restricts CSP(H) to input structures G that are both 1-full and 2-full, i.e., structures G that have each vertex in a unique unary relation (list), and each pair of vertices in a unique binary relation, in a unique way. In particular, it follows that the binary relations on G are antisymmetric. However, in our examples below, we shall have structures H which consist of symmetric binary relations. As a consequence, the orientations of the arcs of the various binary relations on G are irrelevant, and we may view G as also having symmetric binary relations.

We first consider one very small example problem from [FH06$^+$a]. (A related problem was introduced in [CEHS04] and is discussed below.) Let H_3 be the structure with three vertices, $0, 1, 2$, all eight unary relations corresponding to the subsets of $\{0, 1, 2\}$, and three symmetric binary relations S_0, S_1, S_2, where each $S_i = V(H_3) \times V(H_3) \setminus (i, i)$. We shall be mostly interested in the problem $\text{CSP}_{\{1,2\}}(H_3)$. However, it will also be of interest to consider the problem $\text{CSP}_2(H_3^-)$, where H_3^- is obtained from H_3 by deleting all unary relations. It is easy to see that $\text{CSP}_2(H_3^-)$ can be reformulated in purely graph theoretic language; we shall do that in the next subsection.

3.5 Compatible k-colourings

Given a complete graph $G = K_n$ whose edges are coloured by $0, 1, \ldots, k$, a *compatible vertex colouring* of G assigns colours $0, 1, \ldots, k$ to the vertices of G so that no edge $e = uv$ of G has the same colour on e, u, and v.

The problem $\text{CSP}_2(H_3^-)$ introduced above is precisely the problem of existence of a compatible 3-colouring. In other words, *given a complete graph $G = K_n$ with edges coloured $0, 1, 2$, is there a compatible vertex colouring?* In the problem $\text{CSP}_{\{1,2\}}(H_3)$, the vertices v of the input structure G are furthermore equipped with lists $L(v)$ of allowed images.

The problem of compatible 2-colouring has a simple form and a nice solution. Indeed, a complete graph G with edges coloured $0, 1$ may be viewed as just an ordinary graph (by taking the edges of colour 0 to be absent); thus the problem has become one of partitioning the vertices of G into two sets, those coloured 0 – which must form a clique – and those coloured 1 – which must form an independent set. Graphs G which admit such a partition are called *split* graphs [Gol80] and can be recognized in linear time. (In fact G is a split graph if and only if it is chordal and contains no induced $2K_2$ [Gol80].) The addition of lists makes only a very small difference and both linear time algorithms and forbidden induced subgraph characterizations still apply [HKNP04].

The problem of compatible 4-colourings (and k-colourings with $k > 4$) is NP-complete, even without lists. Indeed, given a graph G, it is easy to construct a graph G' with edges coloured $0, 1, 2, 3$, which can be compatibly coloured if and only if G is 3-colourable (in the usual sense). It is enough to let G' consist of two disjoint copies of G with each edge of G replaced by three parallel edges coloured $0, 1, 2$, and colouring all remaining pairs of vertices by 3. Indeed, any 3 colouring of G (repeated on each copy of G in G') is a compatible colouring of G'; and conversely, any compatible colouring of G' must not use colour 3 on the vertices of one copy of G (since they are linked by all edges coloured 3), and hence induce a valid 3-colouring on G. (Note that the graph G' has parallel edges, or equivalently, some pairs of vertices of G are related in several relations. It is not hard to replace these multiple edges by substituting p independent vertices for each vertex of G, and joining two independent sets corresponding to a pair of vertices joined by edges of colours $0, 1, 2$, by a complete bipartite graph with edges coloured $0, 1, 2$, so that between any two subsets of at least $p/2$ vertices there is an edge of each colour $0, 1, 2$. If p is large enough, random colourings have this property.)

The complexity of the compatible 3-colouring problem (with or without lists) is not known. It is not likely to be NP-complete, because of the following result [FH06$^+$a].

Theorem 3.7. *There is an $n^{O(\log n)}$ algorithm to solve the compatible 3-colouring problem, with lists.*

Proof. Initial lists are given, $L(v) \subseteq \{0, 1, 2\}$, for all vertices v of G. (In the problem without lists, we can take all initial $L(v) = \{0, 1, 2\}$.) At each stage, the lists will be reduced, by replacing a problem by a set of subproblems with smaller lists, until we reach a stage when all lists have at most two vertices. Such problems can be solved by a standard application of a two-satisfiability algorithm [APT79]; indeed choosing an image from a list of size two represents a Boolean choice, and all constraints are defined over pairs of vertices, yielding clauses of size two [FHKM03].

Let S denote the (changing) set of vertices v with $L(v) = \{0, 1, 2\}$. We say that i is the *majority colour* at $v \in S$ if it occurs on at least a third of the edges from v to the other vertices of S.

We reduce the current problem to $|S| + 1$ subproblems as follows. In the first subproblem, we avoid giving any vertex $v \in S$ its majority colour. This results in all lists of size at most two, and can be tested in linear time. In the remaining $|S|$ subproblems, we assume for each $v \in S$ in turn, that (at least) v receives its majority colour i; this allows us to remove i from the list of at least $|S|/3$ other vertices.

We obtain the recurrence $T(s) \leq (1 + sT(2s/3))T_2(n)$, where $s = |S|$ and $T_2(n)$ is the time for solving an instance of two-satisfiability on n variables and at most n^2 clauses. It is easy to see that the solution is $T(n) = n^{O(\log n)}$. □

No NP-complete problem is known to have an $n^{O(\log n)}$ algorithm, and the nature of polynomial time reductions ensures that if one NP-complete problem did have such an algorithm, then so would all the others. Thus we take our result as evidence that the compatible 3-colouring problem is not likely to be NP-complete. Yet no polynomial time algorithm is known for the problem. There are better algorithms than the simple example above – the currently best algorithm has complexity $n^{O(\log n/\log\log n)}$ [FHKS05].

3.6 Quasi-dichotomy of General Λ-full Problems

What can one expect of the general problems $\mathrm{CSP}_\Lambda(H)$, if even this first tiny example problem defeated our desire for dichotomy? It turns out that this is as bad as the situation can get. While we clearly cannot prove (and do not conjecture) dichotomy for full constraint satisfaction problems, a weaker version of dichotomy may be possible. Let us say that a problem is *quasi-polynomial*, if it admits an $n^{O(\log n)}$ algorithm [FHKM03]. If each problem from a class $\mathcal{C}$ of problems is NP-complete or quasi-polynomial, we have a new kind of dichotomy; to distinguish it from the usual kind, we shall say that the class $\mathcal{C}$ has *quasi-dichotomy.*

Under a mild restriction, we can ensure that for the most general problems $\mathrm{CSP}_\Lambda(H)$ we have quasi-dichotomy, regardless of the integers in Λ (as long as $1 \in \Lambda$), and of the arities of the relations in H [FH06$^+$a].

We say that H is *three-inclusive* if for any $R \in \mathcal{U}(H)$ and any $R' \subseteq R$ with $|R'| \leq 3$ we also have $R' \in \mathcal{U}(H)$.

Theorem 3.8. *Suppose* $1 \in \Lambda$. *For any three-inclusive structure* H, *the problem* $CSP_\Lambda(H)$ *is NP-complete or quasi-polynomial.*

We may view the theorem as a counterpart to Bulatov's dichotomy theorem for conservative structures, Theorem 3.1, which is crucially used in the proof of Theorem 3.8. While Theorem 3.1 requires H to be conservative (contain all unary relations), Theorem 3.8 only requires H to be three-inclusive. While Theorem 3.1 only restricts the input structures to be 1-full (every vertex has a list), yielding a problem equivalent to a constraint satisfaction problem, Theorem 3.8 allows any restriction to Λ-full input structures (as long as $1 \in \Lambda$), resulting in problems that are not necessarily constraint satisfaction problems $\mathrm{CSP}(H)$. Of course, the price we pay for such generality is that we don't get dichotomy, only quasi-dichotomy.

4 Back to Graph Colourings

We now return to colourings of graphs. We shall focus on structures H which have all unary relations, and two other relations denoted by E and N, both binary and symmetric. To simplify the descriptions we shall also assume that $E \cup N = V(H) \times V(H)$.

Consider an input structure G of the problem $\mathrm{CSP}_{\{1,2\}}(H)$, for such a structure H. Since G is 2-full, it has for each pair of vertices u, v exactly one of the ordered pairs $(u, v), (v, u)$ in exactly one of E or N. In fact, it does not matter whether it is (u, v) or (v, u), since both relations E and N are symmetric, and we shall not worry about the distinction. In other words, we shall consider E and N also to be symmetric relations on G, and interpret 2-fullness to mean that each pair of distinct vertices of G is related either in E or in N. Thus G may be viewed as a graph with edge set E (where N are the non-edges of G, i.e., the edges of the complement $\overline{G}$). Now a homomorphism of the structure G to the structure H is a *full homomorphism* of the graph G to the structure H [HN04b], in the sense that it must preserve both the relations E and N. Thus full homomorphisms to H correspond to partitions (colourings) of the vertices of G into (colour) classes $V_h, h \in V(H)$, in which the part (colour class) V_h is a clique if the loop $hh \in E$, an independent set if $hh \in N$, or unrestricted if $hh \in E \cap N$; and between two parts $V_h, V_{h'}$ we have all edges if $hh' \in E$, no edges if $hh' \in N$, or anything if $hh' \in E \cap N$. Such a partition is called a *H-partition of* G, and we shall refer to $\mathrm{CSP}_2(H)$ as the *H-partition problem* and to $\mathrm{CSP}_{\{1,2\}}(H)$ as the *list H-partition problem.* (See [FH06$^+$a, FH06$^+$b, FH04, FHKM03, FHKNP05, FHKS05, FHK04, FHH05, HKNP04, HN04b] where these problems are introduced via matrix notation; the matrices involved may be viewed as the adjacency matrices of the structures H.)

This introduces a novel perspective on colourings. In Figure 2 we show some example structures H of this type. The edges of E are depicted as solid lines, the edges of N as dotted lines. (The unary relations are not depicted, but recall that the structures H are assumed to be conservative, i.e., to include all unary relations.)

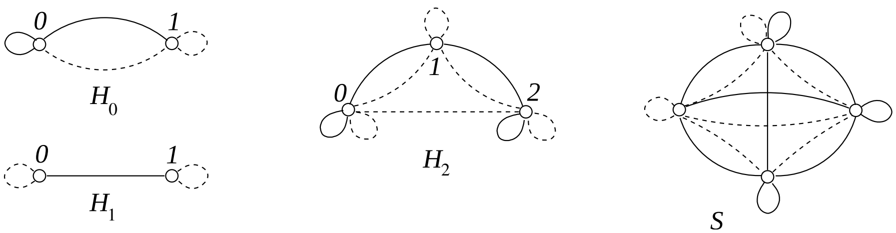


Fig. 2. Some example structures

A graph G admits a full homomorphism to H_0 if and only if it is a split graph. The vertices mapped to 0 form an independent set and those that map to 1 form a clique. Thus the H_0-partition problem is precisely the problem of recognizing split graphs, and the list H_0-partition problem is its list version; both are solvable in linear time as mentioned earlier [Gol80, HKNP04].

A graph G admits a full homomorphism to H_1 if and only if it is complete bipartite; it is a simple exercise to show that this is the case if and only if G does not have an induced $K_1 \cup K_2$.

A graph G admits a surjective full homomorphism to H_2 if and only if it admits a clique cutset. (A full homomorphism f of G to H is surjective if $f(V(G)) = V(H)$.) Indeed, the (nonempty) set of vertices that map to 1 forms a clique and there are no edges joining the (nonempty) sets of vertices that map to 0 and to 2. The problem of existence of a clique cutset is solvable in linear time [Tar85], as is its list version [FHKM03, HN04b]. Clique cutsets play an important role in the study of chordal graphs [Gol80, Tar85].

Many other concepts that occur in the study of graph perfection can be expressed in terms of full homomorphisms, surjective full homomorphisms, or list full homomorphisms, to various structures H; these include the concepts of *homogeneous set, skew cutset, and joins* of various kinds [CRST06$^+$, Chv85, Gol80], cf. [FHKM03]. As an application of Theorem 3.8, we see that all list H-partition problems are quasi-polynomial or NP-complete. A finer classification, as NP-complete or polynomial time solvable, of all list H-partition problems for structures H with up to four vertices (up to three vertices in the case when E and N are not assumed symmetric), has been undertaken in [CEHS04, FHKM03, FHK04, FKKB00]. All the quasi-polynomial problems have in fact been shown to be polynomial time solvable in these cases – with one exception, namely for the list S-partition problem, where S is the last structure in Figure 2. This problem is open, and has been dubbed the 'stubborn problem' by [CEHS04]; it is a relative of the problem $\mathrm{CSP}_{\{1,2\}}(H_3)$ discussed earlier, and we have for it the same bound of $n^{O(\log n/\log\log n)}$ [FHKS05].

A particularly nice situation arises when H itself is 2-full, i.e., each pair of vertices $u, v \in V(H)$ is related either in E or in N. This is typified by the structure H_1 in Figure 2. It is shown in [FH04] that for each such H there is a forbidden induced subgraph characterization of the existence of full homomorphisms to H:

Theorem 4.1. *If H is itself 2-full, then there exists a finite set of graphs $G_1, G_2, \ldots, G_p$ such that a graph G admits a full homomorphism to H if and only if it does not have any G_i as an induced subgraph.*

We say that $G_1, G_2, \ldots, G_p$ are the *forbidden induced subgraphs* for H. Since each loop vv of H is either in E or in N, we denote by k the number of loops in E and $\ell = n - k$ the number of loops in N. In fact, we proved in [FH04] that all minimal forbidden induced subgraphs have at most $(k+1)(\ell+1)$ vertices, and that there are *at most two* minimal forbidden induced subgraphs with exactly $(k+1)(\ell+1)$ vertices. For instance in the structure H_1 in Figure 2, $k = 0$, $\ell = 2$ and the unique minimal forbidden induced subgraph $K_1 \cup K_2$ has exactly $(k+1)(\ell+1) = 3$ vertices.

When H is not assumed to be 2-full, we may distinguish between vertex pairs u, v related in both E and N (in other words with $uv \in E \cap N$), which we shall call *weak pairs*, and pairs u, v related in only one of E, N (in other words with $uv \in E - N \cup N - E$), which we shall call *strong pairs*. (Recall that we assume that $E \cup N = V(H) \times V(H)$, so that each pair is either weak or strong.) Thus a 2-full structure has all pairs u, v strong, including pairs with $u = v$. Note that it makes sense to assume that pairs vv are always strong – at least for H-partition problem without lists. (If vv is weak, all vertices of an input graph G can be mapped to v.) In this situation, the possibility of general characterizations by a finite set of forbidden induced subgraphs is investigated in [Xie06].

A number of related results have been obtained for special classes of graphs [FH06$^+$b, FHKNP05, FHH05, HKNP04]. One might ask, for instance, what happens if these problems are restricted to the class of *perfect* input graphs G. It turns out that there are structures H for which $\mathrm{CSP}_2(H)$ (without lists) is NP-complete when restricted to perfect (or even chordal) graphs G [FHKNP05]. Nevertheless, there are large classes of natural structures H for which the (list) H-partition problem is polynomial time solvable for perfect graphs G, and in many cases can be described by a finite set of forbidden induced subgraphs. Let V_E denote the set of vertices v with the loop in E and V_N the set of those with a loop in N. We say that H is *normal*, if either all pairs of vertices $u, v \in V_E$ are weak, or all are strong; all pairs of vertices $u, v \in V_N$ are weak, or all are strong; and all pairs of vertices $u \in V_E, v \in V_N$ are weak, or all are strong. For future reference, we shall call the last kind of pairs u, v *crossed*. For normal structures H, the existence of a full homomorphism to H for perfect graphs can be characterized by a finite set of forbidden induced subgraphs.

Theorem 4.2. *If H is a normal structure, then there exists a finite set of graphs $G_1, G_2, \ldots, G_p$, such that a perfect graph G admits a full homomorphism to H if and only if it does not have any G_i as an induced subgraph.*

In general, the bounds on the size of G_i are not polynomial in k, ℓ; however, in certain cases the bound $(k + 1)(\ell + 1)$ does apply [FH06$^+$b].

For chordal graphs, we now consider the special normal structures H in which all pairs of distinct vertices are weak. In this case, there turns out to be just one forbidden induced subgraph, with $(k + 1)(\ell + 1)$ vertices, namely the union of $(\ell + 1)$ cliques of size $(k + 1)$ [HKNP04]. A linear-time algorithm to solve each such H-partition problem for chordal graphs is described in [HKNP06$^+$], cf. [HKNP04]. It can be viewed as a common generalization of the algorithms for computing the chromatic number on chordal graphs, and on complements of chordal graphs [Gol80].

We note that there are structures H with polynomial time solvable H-partition problems for which the existence of full homomorphism to H cannot be characterized by a finite set of forbidden induced subgraphs; for instance the two-vertex structure H for which $\mathrm{CSP}(H)$ is the problem of two-colourability. Similar phenomena occur also for problems $\mathrm{CSP}(H)$ restricted to perfect graphs [FH06$^+$b].

Returning to chordal graphs G, can extend the class of structures H with polynomial time H-partition algorithms [FHKNP05]. We say that a structure H is *crossed* if the crossed pairs u, v with $u \in V_E, v \in V_N$ have the following property: some vertices of H are only incident with crossed pairs that are strong, and all crossed pairs not incident with one of these vertices are weak. Note that a normal structure has all crossed pairs strong (all vertices of H are incident with strong crossed pairs) or all crossed pairs weka (no vertices of H have this property). Thus crossed structures substantially generalize normal structures, and we have the following result [FHKNP05].

Theorem 4.3. *If H is crossed, then the list H-partition problem for chordal graphs can be solved in polynomial time.*

On the other hand, as mentioned earlier, there are structures H for which the H-partition problems are NP-complete (even without lists) for chordal graphs [FHKNP05]. In fact, the Dichotomy Conjecture is hiding amongst the H-partition problems, even for perfect graphs.

Theorem 4.4. *Every constraint satisfaction problem CSP(H) is polynomial time equivalent to some H'-partition problem, restricted to perfect graphs G.*

Finally, we close with the following good news. There is a natural class of graphs, namely the class of *cographs*, for which all (list) H-partition problems are polynomial time solvable, and characterized by a finite set of forbidden induced subgraphs [FHH05].

Theorem 4.5. *For every structure H, the existence of a (list) full homomorphism to H from an input cograph G can be characterized by a finite set of forbidden induced subgraphs.*

For bounds on the size of these forbidden induced subgraphs we refer the interested reader to [FHH05].

The author gratefully acknowledges suggestions and advice from A. Bulatov and B. Larose.

References

[APT79] B. Aspvall, F. Plass and R. E. Tarjan, A linear time algorithm for testing the truth of certain quantified Boolean formulas, *Information Processing Letters*, Vol. 8 (1979), 121–123.

[BDS87] H.-J. Bandelt, A. Dählmann, H. Schütte, Absolute retracts of bipartite graphs, *Discrete Appl. Math.* 16 (1987), 191–215.

[BFH93] H.-J. Bandelt, M. Farber, and P. Hell, Absolute reflexive retracts and absolute bipartite retracts, *Discrete Applied Math.* 44 (1993), 9–20.

[BP91] H.-J. Bandelt, E. Pesch, Efficient characterizations of n-chromatic absolute retracts, *J. Combinatorial Theory B* 53 (1991), 5–31.

[BH90] J. Bang-Jensen and P. Hell, The effect of two cyles on the complexity of colourings by directed graphs, *Discrete Applied Math.* 26 (1990), 1–23.

[BHM88] J. Bang-Jensen, P. Hell, and G. MacGillivray, The complexity of colouring by semicomplete digraphs, *SIAM J. on Discrete Math.* 1 (1988), 281–298.

[BHM92] J. Bang-Jensen, P. Hell, and G. MacGillivray, On the complexity of colouring by superdigraphs of bipartite graphs, *Discrete Math.* 109 (1992), 27–44.

[BHM95] J. Bang-Jensen, P. Hell, and G. MacGillivray, Hereditarily hard H-colouring problems, *Discrete Math.* 138 (1995), 75–92.

[BHT92] M. Birc, M. Hujter, and Zs. Tuza, Precoloring extension I; interval graphs, *Discrete Math.* 100 (1992), 267–279.

[BFHHM04] R. Brewster, T. Feder, P. Hell, J. Huang, and G. MacGillivray, Near-unanimity functions and varieties of graphs, manuscript 2004.

[Bor31] K. Borsuk, Sur les rétractes, *Fund. Math.* 17 (1931), 152–170.

[BHLL02] A. Branstädt, P. Hammer, V. Bang Le, and V. Lozin, Bisplit graphs, DIMACS Technical Report 2002–44 (2002).

[Bul03] A. A. Bulatov, Tractable conservative constraint satisfaction problems, In *Proceedings of the 18th IEEE Annual Symposium on Logic in Computer Science (LICS 2003)*, 321–330.

[Bul02a] A. A. Bulatov, Maltsev constraints are tractable, Technical report PRG-RR-02-05 Oxford University (2002).

[Bul02b] A. A. Bulatov, A dichotomy constraint on a three-element set, In *Proceedings of the 43rd IEEE Symposium on Theory of Computing* (2002), 649–658.

[Bul06$^+$] A. A. Bulatov, H-colouring dichotomy revisited, *Theoretical Computer Science*, to appear.

[BJ01] A. A. Bulatov and P. Jeavons, Algebraic structures in combinatorial problems, Technical Report MATH-AL-4-2001m Technische universität Dresden, Germany, 2001.

[BJK05] A. A. Bulatov, P. Jeavons, and A. A. Krokhin, Classifying complexity of constraints using finite algebras, *SIAM J. on Computing* 34 (2005), 720–742.

[CEHS04] K. Cameron, E. E. Eschen, C. T. Hoang and R. Sritharan, The list partition problem for graphs, *Proc. 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) 2004*, 391–399.

[CKKLS04] E. Cheng, R. P. Kleinberg, S. G. Kruk, W. A. Lindsey, and D. E. Steffy, A strictly combinatorial approach to a university exam scheduling problem, *Congr. Numer.* 167 (2004) 121–132.

[CRST06$^+$] M. Chudnovsky, N. Robertson, P. Seymour, and R. Thomas, The strong perfect graph theorem, to appear in the *Annals of Math.*

[Chv85] V. Chvátal, Star-cutsets and perfect graphs, *J. Combinatorial Theory B* 39 (1985), 189–199.

[CCJ94] D. Cohen, M. Cooper, and P. Jeavons, Characterizing tractable constraints, *Artificial Intelligence* 65 (1994), 347–361.

[CKS01] N. Creignou, S. Khanna, and M. Sudan, *Complexity Classifications of Boolean Constraint Satisfaction Problems*, *SIAM Monographs on Discrete Math. and Applications*, vol. 7 (2001).

[Dal00] V. Dalmau, A new tractable class of constraint satisfaction problems, in *Proceedings 6th International Symposium on Artificial Intelligence and Mathematics*, 2000.

[DF03] V. Dalmau and D. Ford, Generalized satisfiability with k occurrences per variable: a study through delta-matroid parity, *Mathematical Foundations of Computer Science* (MFCS 2003), *Lecture Notes in Computer Science* 2747, Springer 2003, 358–367.

[Dec92] R. Dechter, Constraint networks, In *Encyclopedia of Artificial Intelligence*, 1992, 276–285.

[DG84] W. F. Dowling and J. H. Gallier, Linear-time algorithms for testing the satisfiability of propositional Horn formulae, *J. Logic Programming* 1 (1984) 267–284.

[Fed01a] T. Feder, Fanout limitations on constraint systems, *Theoretical Computer Science* 255 (2001), 281–293.

[Fed01b] T. Feder, Homomorphisms to oriented cycles and k-partite satisfiability, *SIAM J. Discrete Math.* 14 (2001), 471–480.

[Fed04] T. Feder, Constraint satisfaction: a personal perspective, manuscript 2004.

[Fed06$^+$] T. Feder, A dichotomy theorem on fixed points of several nonexpansive mappings, to appear in *SIAM J. Discrete Math.*

[FF06$^+$] T. Feder and D. Ford, Classification of bipartite Boolean constraint satisfaction through delta-matroid intersection, manuscript.

[FH98] T. Feder and P. Hell, List homomorphisms to reflexive graphs, *J. Comb. Theory Series B* 72 (1998), 236–250.

[FH06$^+$a] T. Feder and P. Hell, Full constraint satisfaction problems, to appear in *SIAM J. on Computing.*

[FH06$^+$b] T. Feder and P. Hell, Matrix partitions of perfect graphs, to appear in *Berge Special Issue of Discrete Math.*

[FH04] T. Feder and P. Hell, On realizations of point determining graphs, and obstructions to full homomorphisms, manuscript 2004.

[FH03] T. Feder and P. Hell, The complexity of retraction problems for reflexive digraphs, manuscript 2003.

[FHH99a] T. Feder, P. Hell, and J. Huang, List homomorphisms and circular arc graphs, *Combinatorica* 19 (1999), 487–505.

[FHH99b] T. Feder, P. Hell, and J. Huang, Bi-arc graphs and the complexity of list homomorphisms, *J. Graph Theory* 42 (1999), 61–80.

[FHH06$^+$] T. Feder, P. Hell, and J. Huang, List homomorphisms of graphs with bounded degrees, to appear in *Discrete Math.*

[FHH04] T. Feder, P. Hell, and J. Huang, List homomorphisms to reflexive digraphs, manuscript 2004.

[FHKM03] T. Feder, P. Hell, S. Klein, and R. Motwani, Complexity of list partitions, SIAM J. Discrete Mathematics 16 (2003), 449–478.

[FHKNP05] T. Feder, P. Hell, S. Klein, L. Nogueira, and F. Protti, List matrix partitions of chordal graphs, *Theoretical Computer Science* 349 (2005), 52–66.

[FHKS05] T. Feder, P. Hell, D. Král', and J. Sgall, Two algorithms for list matrix partitions, *Proc. 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)* 2005, 870–876.

[FHM03] T. Feder, P. Hell, and B. Mohar, Acyclic homomorphisms and circular colorings of digraphs, *SIAM J. Discrete Math.* 17 (2003), 161–169.

[FHK04] T. Feder, P. Hell, and K. Tucker-Nally, Digraph matrix partitions and trigraph homomorphisms, manuscript 2004.

[FHH05] T. Feder, P. Hell, and W. Hochstättler, Generalized colourings (matrix partitions) of cographs, *GT04 Special Berge Volume*, to be published by Birkhauser.

[FMS04] T. Feder, F. Madelaine, and I. A. Stewart, Dichotomies for classes of homomorphism problems involving unary functions, *Theoretical Computer Science* 314 (2004), 1–43.

[FV98] T. Feder and M. Y. Vardi, The computational structure of monotone monadic SNP and constraint satisfaction: a study through Datalog and group theory, *SIAM J. Comput.* 28 (1998), 236–250, (ALSO in *STOC* 25 (1993), 612–622).

[FKKB00] C. M. H. de Figueiredo, S. Klein, Y. Kohayakawa and B. Reed, Finding skew partitions efficiently, *J. Algorithms* 37 (2000), 505–521.

[GHN00] A. Galluccio, P. Hell, J. Nešetřil, The complexity of H-colouring of bounded degree graphs, *Discrete Math.* 222 (2000), 101–109.

[GJ79] M. Garey and D. S. Johnson, *Computers and Intractability: a Guide to the Theory of NP-completeness*, Freeman and Company, 1979.

[Gol80] M. C. Golumbic, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, New York, 1980.

[GWW92] W. Gutjahr, E. Welzl and G. Woeginger, Polynomial graph-colorings, *Discrete Applied Math* 35 (1992), 29–45.

[Hell72] P. Hell, *Rétractions des graphes*, PhD thesis, Université de Montréal, 1972.

[Hell74a] P. Hell, Absolute planar retracts and the four color conjecture, *J. Combinatorial Theory B* 17 (1974), 5–10.

[Hell74b] P. Hell, Absolute retracts in graphs, in *Graphs and Combinatorics* (R. A. Bari, F. Harary, eds.), Springer-Verlag Lecture Notes in Mathematics 406 (1974), 291–301.

[HKNP04] P. Hell, S. Klein, L. Tito Nogueira, and F. Protti, Partitioning chordal graphs into independent sets and cliques, *Discrete Applied Math.* 141 (2004), 185–194.

[HKNP06+] P. Hell, S. Klein, L. Tito Nogueira, and F. Protti, Packing r-cliques in chordal graphs, *Annals of O.R.* 138 (2005), 179–187.

[HN92] P. Hell, J. Nešetřil, The core of a graph, *Discrete Math.* 109 (1992), 117–126.

[HN90] P. Hell and J. Nešetřil, On the complexity of H-coloring, *J. Comb. Theory, Series B* 48 (1990), 92–110.

[HN04a] P. Hell and J. Nešetřil, Counting list homomorphisms and graphs with bounded degrees, in *Graphs, Morphisms and Statistical Physics* (J. Nešetřil and P. Winkler, eds.) *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Volume 63 (2004), 105–112.

[HN04b] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Oxford University Press 2004.

[HNZ96] P. Hell, J. Nešetřil, and X. Zhu, Complexity of tree homomorphisms, *Discrete Applied Math.* 70 (1996), 23–36.

[HR87] P. Hell, I. Rival, Absolute retracts and varieties of reflexive graphs, *Canad. J. Math.* 39 (1987), 544–567.

[Horn51] A. Horn, On sentences which are true of direct unions of algebras, *J. Symbolic Logic* 16 (1951), 14–21.

[HT93] M. Hujter and Zs. Tuza, Precoloring extension II; graph classes related to bipartite graphs, *Acta Math. Universitatis Comenianae* 62 (1993), 1–11.

[HT96] M. Hujter and Zs. Tuza, Precoloring extension III; classes of perfect graphs, *Combin. Probab. Comput.* 5 (1996), 35–56.

[JMP86] E. M. Jawhari, D. Misane, M. Pouzet, Retracts: graphs and ordered sets from the metric point of view, *Contemp. Math. (Amer. Math. Soc.)* 57 (1986), 175–226.

[Jea98] P. Jeavons, On the algebraic structure of combinatorial problems, *Theoretical Computer Science* 200 (1998), 185–204.

[JCG97] P. Jeavons, D. Cohen, and M. Gyssens, Closure properties of constraints, *J. of the ACM* 44 (1997), 527–548.

[JCG95] P. Jeavons, D. Cohen, and M. Gyssens, A unifying framework for tractable constraints, *Proc. 1st International Conference on Constraint Programming, CP'95*, Cassis 1995.

[Joh82] D. S. Johnson, The NP-completeness columns: an ongoing guide, *J. Algorithms* 3 (1982), 89–99.

[Kum92] V. Kumar, Algorithms for constraint-satisfaction problems, *AI Magazine* 13 (1992), 32–44.

[Lad75] R. E. Ladner, On the structure of polynomial time reducibility, *J. Assoc. Comput. Mach.* 22 (1975), 155–171.

[Lar05] B. Larose, Taylor operations on finite reflexive structures, manuscript 2005.

[LT02] B. Larose and C. Tardif, Projectivity and independent sets in powers of graphs, J. Graph Theory 40 (2002), 162–171.

[LZ04] B. Larose and L. Zádori, Finite posets and topological spaces in locally finite varieties, *Algebra Universalis* 52 (2004), 119–136.

[LZ03] B. Larose and L. Zádori, The complexity of the extendibility problem for finite posets, *SIAM J. on Discrete Math.* 17 (2003), 114–121.

[LM92] P. Lincoln and J. C. Mitchell, Algorithmic aspects of type inference with subtypes, in *Conf. Rec. 19th ACM Symp. on Principles of Programming Languages* (1992), 293–304.

[LS86] V. Lotfi and S. Sarin, A graph coloring algorithm for large scale scheduling problems, *Comput. Oper. Res.* 13 (1986), 27–32.

[ŁN03a] T. Łuczak, J, Nešetřil, A note on projective graphs, *ITI series* 2003, No. 130.

[ŁN03b] T. Łuczak, J. Nešetřil, Towards probabilistic analysis of the dichotomy problem, *KAM-DIMATIA Series* 2003-640, Charles Univ. Prague.

[Marx03] D. Marx, Precoloring extension on chordal graphs, manuscript 2003.

[Mes89] P. Meseguer, Constraint satisfaction problem: an overview, *AICOM* 2 (1989), 3–16.

[Mit84] J. C. Mitchell, Coercion and type inference (summary), in *Conf. Rec. 11th ACM Symp. on Principles of Programming Languages* (1984), 175–185.

[Mon74] U. Montanari, Networks of constraints: Fundamental properties and applications to picture processing, *Information Sciences* 7 (1974), 95–132.

[Pes88] E. Pesch, *Retracts of Graphs*, Athenaeum Verlag, Frankfurt 1988.

[PT96] V. Pratt and J. Tiuryn, Satisfiability of inequalities in a poset, *Fundamenta Informaticae* 28 (1996), 165–182.

[PT80] A. Pultr, V. Trnková, *Combinatorial, Algebraic and Topological Representations of Groups, Semigroups and Categories*, North-Holland, Amsterdam, 1980.

[Qui85] A. Quilliot, A retraction problem in graph theory, *Discrete Math.* 54 (1985), 61–71.

[SO74] A. Salazar and R. V. Oakford, A graph formulation of a school scheduling algorithm, *Comm. ACM* 17 (1974), 696–698.

[Sch78] T. J. Schaefer, The complexity of satisfiability problems, *Proc. 10th ACM Symp. on Theory of Computing* (1978), 216–226.

[Tar85] R. E. Tarjan, Decomposition by clique separators, *Discrete Mathematics* 55 (1985), 221–232.

[Tsa93] E. P. K. Tsang, *Foundations of Constraint Satisfaction*, Academic Press, London and San Diego (1993).

[Var00] M. Y. Vardi, Constraint satisfaction and database theory: a tutorial, *Proceedings of the 19th Symposium on Principles of Database Systems (PODS 2000)*, 76–85.

[Vik03] N. Vikas, Computational complexity of compaction to reflexive cycles *SIAM J. on Computing* 32 (2003), 253–280.

[WO89] M. Wand and P. M. O'Keefe, On the complexity of type inference with coercion, in *Conf. on Functional Programming Languages and Computer Architecture* (1989), 293–298.

[Xie06] W. Xie, *Forbidden subgraph characterizations of matrix partitions*, M.Sc. thesis, Simon Fraser University 2006.

Part V

Graph Colorings

Thresholds for Path Colorings of Planar Graphs

Glenn G. Chappell[1], John Gimbel[2], and Chris Hartman[3]

[1] Department of Computer Science, University of Alaska,
Fairbanks, AK 99775-6670 USA
chappellg@member.ams.org

[2] Department of Mathematics and Statistics, University of Alaska,
Fairbanks, AK 99775-6660 USA
ffjgg@uaf.edu

[3] Department of Computer Science, University of Alaska,
Fairbanks, AK 99775-6670 USA
hartman@cs.uaf.edu

Summary. A graph is path k-colorable if it has a vertex k-coloring in which the subgraph induced by each color class is a disjoint union of paths. A graph is path k-choosable if, whenever each vertex is assigned a list of k colors, such a coloring exists in which each vertex receives a color from its list.

It is known that every planar graph is path 3-colorable [Poh90, God91] and, in fact, path 3-choosable [Har97]. We investigate which planar graphs are path 2-colorable or path 2-choosable. We seek results of a "threshold" nature: on one side of a threshold, every graph is path 2-choosable, and there is a fast coloring algorithm; on the other side, determining even path 2-colorability is NP-complete.

We first consider maximum degree. We show that every planar graph with maximum degree at most 4 is path 2-choosable, while for $k \geq 5$ it is NP-complete to determine whether a planar graph with maximum degree k is path 2-colorable.

Next we consider girth. We show that every planar graph with girth at least 6 is path 2-choosable, while for $k \leq 4$ it is NP-complete to determine whether a planar graph with girth k is path 2-colorable. The case of girth 5 remains open.

AMS Subject Classification. Primary: 05C15, Secondary: 05C38, 05C85.
Keywords. Path coloring, list coloring, planar graph, girth, NP-complete.

1 Introduction

Graphs will be finite and, unless stated otherwise, simple and undirected. For undefined terms and concepts the reader is referred to [CL96].

Definition 1.1. *A* path coloring *of a graph G is a vertex coloring of G so that each color class induces a linear forest, that is, a disjoint union of paths.*

A path k-coloring *is a path coloring using at most k colors. If G has a path k-coloring, we say that G is* path k-colorable. □

Clearly, if a graph has a proper k-coloring, then it is path k-colorable. Thus, by the Four-Color Theorem [AH77, AHK77, RSST97], every planar graph is path 4-colorable. Poh [Poh90] and Goddard [God91] independently showed that, in fact, 3 colors suffice, thus verifying a conjecture of Akiyama, Era, Gervacio, and Watanabe [AEGW89].

Theorem 1.2 (Poh 1990 [Poh90], Goddard 1991 [God91]). *Let G be a planar graph. Then G is path 3-colorable.* □

This result is best-possible; there are planar graphs that are not path 2-colorable. The minimum order of such a graph is 9; an example is shown in Figure 1. By checking all the maximal planar graphs of order 8, one can show that every planar graph of order 8 or less is path 2-colorable. (This corrects an oversight in [AEGW89], in which a figure purports to show an 8-vertex graph with no path 2-coloring.)

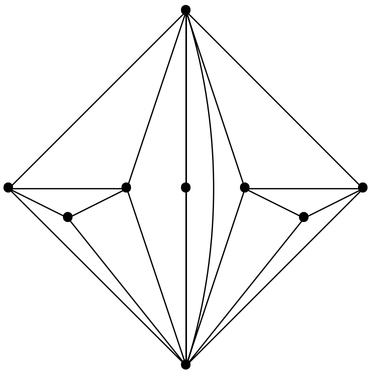


Fig. 1. A planar graph that is not path 2-colorable. This graph has the minimum order, 9, for a graph with this property

A 1-*defective coloring* of a graph G is a vertex coloring of G in which each color class induces a graph with maximum degree at most 1. We denote the least number of colors in a 1-defective coloring of G by $\chi_1(G)$. For a discussion of this parameter the reader is referred to [AJ85, CGJ97, CCW86]. By [Cow93], it is NP-complete to determine whether $\chi_1(G) \leq 2$, for graphs in general. Given G, let us form G' as follows. For each vertex v of G, add three new vertices, which form a K_3. Join v to each of these vertices. Then G' is path 2-colorable if and only if $\chi_1(G) \leq 2$. Thus, it is NP-complete to determine path 2-colorability, for graphs in general.

Suppose we are given a graph G and an integer m. For each vertex v of G let us create two disjoint copies of K_{2m-1} and join these to v. This new graph can be path-colored with m colors if and only if G can be properly colored with m colors. Furthermore, for any fixed $m \geq 3$ it is NP-complete to determine whether $\chi(G) \leq m$. Hence the following.

Remark 1.3. For a fixed $m \geq 2$, it is NP-complete to determine whether a graph is path m-colorable. □

This remark also follows from a result of Farrugia [Far04, Thm. 2].

We investigate which planar graphs are path 2-colorable. It will be useful to define a list-coloring variant of path coloring. We say a graph G is *path k-choosable* if, whenever lists of k colors are assigned to the vertices of G, there exists a path coloring in which each vertex receives a color from its list. Hartman [Har97] showed that every planar graph is path 3-choosable. Of course, any graph that is not path k-colorable is also not path k-choosable (make the lists all the same), and so there exist planar graphs that are not path 2-choosable.

We seek results of a "threshold" nature: ideally, on one side of a threshold, every graph is path 2-choosable, and there is a fast coloring algorithm; on the other side, determining even path 2-colorability is NP-complete.

We first consider maximum degree. We show that every planar graph with maximum degree at most 4 is path 2-choosable, and we present a fast coloring algorithm for some of these graphs. On the other hand, for $k \geq 5$ it is NP-complete to determine whether a planar graph with maximum degree k is path 2-colorable.

Next we consider girth. We show that every planar graph with girth at least 6 is path 2-choosable, and we discuss a fast coloring algorithm. For $k \leq 4$ it is NP-complete to determine whether a planar graph with girth k is path 2-colorable. The case of girth 5 remains open.

2 Maximum Degree

We consider how the maximum degree of a planar graph affects path 2-colorability and choosability. We will make use of the following result of Borodin, Kostochka, and Toft [BKT00, Cor. 5′]. Given a positive integer t, Borodin *et al.* define a graph to be *strictly t-degenerate* if every subgraph contains a vertex that has degree strictly less than t in the subgraph. Thus, for example, by this definition the strictly 2-degenerate graphs are precisely the forests.

Theorem 2.1 (Borodin, Kostochka, Toft 2000 [BKT00]). *Let G be connected, not a complete graph. Let s, t be integers so that $s \geq 2$, $st \geq 3$, and $st \geq \Delta(G)$. Suppose that a list of at least s colors is assigned to each vertex of G. Then there is a vertex coloring of G in which each vertex is assigned a color from its list, and each color class induces a strictly t-degenerate subgraph of maximum degree not greater than t.* □

Using the above result, we can easily prove the following.

Theorem 2.2. *Let k be a positive integer, and let G be a graph with maximum degree at most $2k$. If G has no component isomorphic to K_{2k+1} (or, in the case $k = 1$, a cycle), then G is path k-choosable.*

Proof. If $k \geq 2$, then set $s = k$ and $t = 2$, and apply Theorem 2.1 to each component of G. If $k = 1$, then the conditions of the theorem imply that G is a linear forest, and so every vertex coloring is a path coloring. □

We would like to show that we can find the coloring of Theorem 2.2 with a polynomial-time algorithm. However, we have been able to prove this only when there is no $2k$-regular component.

Theorem 2.3. *Let k be a positive integer. There is a polynomial-time algorithm that, given a graph G that meets the conditions of Theorem 2.2 and has no $2k$-regular component, along with an assignment of lists of size k to the vertices, returns a path coloring of G in which each vertex receives a color from its list.*

Proof. We outline a recursive algorithm. Partition G into components. For each component H, perform the following procedure.

Find a vertex v in H with degree less than $2k$ (this must exist, since H is not $2k$-regular). Remove v and its list from H, and note that the resulting graph and list assignment meets the conditions to be input for the algorithm. Recursively run the algorithm with this input. Replace v and its list in H. Now all of H is colored, except for v. There are k colors in the list assigned to v, and v has degree at most $2k - 1$. Thus, there is a color in the list that is used on at most one neighbor of v. Give v this color.

We have now colored every vertex with a color from its list. However, it is possible that our coloring is not a path coloring. If it is not, we will fix it.

Since the colors on all vertices except v were assigned by the recursive call to the algorithm, removing v results in a path coloring. Since v has degree at most 1 in its color class, returning v to the graph does not create any monochromatic cycles. However the neighbor (if any) of v in its color class may have degree greater than 2 in its color class. We refer to a vertex with degree greater than 2 in its color class as a *bad vertex*. If there is no bad vertex, then we are done. If there is a bad vertex x, then x has degree at least 3 in its color class; x also has degree at most $2k$ in H and a list of k colors. Since there are at most $2k - 3$ neighbors of x that are not in its color class, there must be at least one color in x's list, not used on x, that is used on at most one of x's neighbors. Switch x to this color.

Doing this may make one of x's neighbors into a bad vertex. However, the number of bad vertices does not increase (since x is no longer bad), no monochromatic cycles are created, and the number of edges whose endpoints have the same color is reduced. Thus, this process must eventually end, with no bad vertex being created. When this happens, we have the desired coloring. We return this coloring, ending the algorithm.

This ends the description of the algorithm; we now consider efficiency.

Say the original graph G has n vertices and m edges. For fixed k, m is $O(n)$, since $m < kn$. Each call to the algorithm first finds a vertex with degree less than $2k$. This may require looking at $O(n)$ vertices. Then a single recursive call is made. After the recursive call, the algorithm fixes the coloring using a process that successively reduces the number of edges whose endpoints have the same color. This process thus may require $O(m)$ steps, that is, $O(n)$ steps. Thus, aside from the recursive call, the algorithm requires $O(n)$ steps.

Each call to the algorithm makes at most one recursive call, and each recursive call removes one vertex. The recursion depth is thus $O(n)$, and so the order of the algorithm is $O(n^2)$, which is polynomial-time. □

The more general case, allowing graphs with $2k$-regular components, remains open.

Question 2.4. Can Theorem 2.3 be generalized to include graphs with $2k$-regular components (other than K_{2k+1}, of course)? □

Setting $k = 2$ in Theorems 2.2 and 2.3, and noting that K_5 is not planar, we obtain the first of our desired results on planar graphs.

Corollary 2.5. *If G is a planar graph with maximum degree at most 4, then G is path 2-choosable.* □

Corollary 2.6. *There exists a polynomial-time algorithm that, given a planar graph G with maximum degree at most 4, and no 4-regular component, along with an assignment of lists of size 2 to the vertices, returns a path coloring of G in which each vertex receives a color from its list.* □

On the other hand, we can show the following.

Proposition 2.7. *There exists a planar graph with maximum degree 5 that has no path 2-coloring.* □

Rather than verify the above proposition directly, we prove the following more general result.

Theorem 2.8. *It is NP-complete to determine whether a planar graph with maximum degree 5 is path 2-colorable.*

Proof. We shall see that the 3-SAT problem reduces in polynomial time to this one. We will prove this using a series of configurations (pictured in Figures 2–7) which effectively allow us to perform logical operations using path 2-coloring of planar graphs. Given an instance of 3-SAT, we use these configurations to construct, in polynomial time, a planar graph of maximum degree 5 that is path 2-colorable if and only if the given instance of 3-SAT is satisfiable.

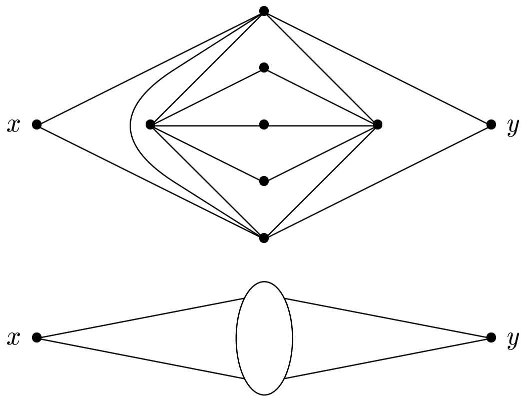


Fig. 2. A planar graph with maximum degree 5, along with a symbol representing it. Path 2-coloring this graph forces x and y to receive the same color. This is used in the proof of Theorem 2.8

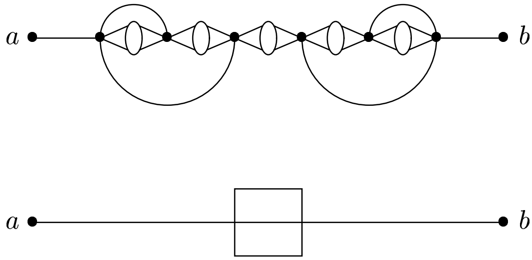


Fig. 3. The "extender", used the proof of Theorem 2.8, along with a symbol representing it. Path 2-coloring this graph forces a and b to receive the same color. Furthermore, both a and b have degree zero in their color classes

Consider the graph at the top of Figure 2, which is represented by the diagram at the bottom of Figure 2. In any path 2-coloring of this graph, vertices x and y must have the same color. Let us use this to build the gadget at the top of Figure 3. This graph is planar and has maximum degree 5. Furthermore, in any path 2-coloring vertices a and b must receive the same color. Let us denote this gadget with the diagram at the bottom of Figure 3 and refer to it as an *extender*. Note that, in every path 2-coloring of the extender, each of vertices a and b receives a color different from that of its neighbor.

We shall refer to the graph at the top of Figure 4 as a *negator* and denote it with the diagram at the bottom of Figure 4. Note that, in any path 2-coloring of the negator, vertices z and w are given different colors. Furthermore, in every path 2-coloring of the negator, each of vertices z and w receives a color different from that of its neighbor.

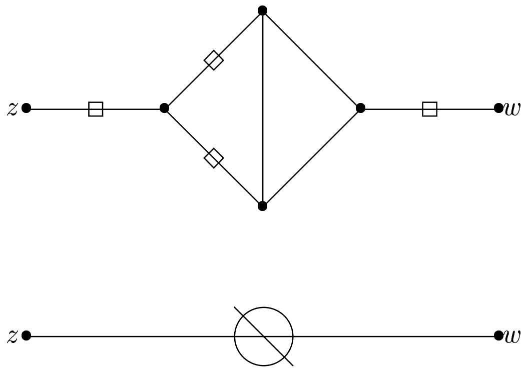


Fig. 4. The "negator", used the proof of Theorem 2.8, along with a symbol representing it. Path 2-coloring this graph forces z and w to receive different colors. Furthermore, both z and w have degree zero in their color classes

Now we consider the graph in Figure 5, which we refer to as the *conjugator*. Suppose this is path 2-colored so that x', y', and z' are all given a color different from that of the vertex labeled T. Then w must receive a color different from that of T as well. Suppose instead that the conjugator is path 2-colored so that at least one of x', y', or z' is given the same color as T. Then w and T must receive this same color. Thus, identifying the color assigned to T with **true** and the other color with **false**, path 2-coloring the conjugator forces w to be colored with the logical-or of the colors of x', y', z'.

Let us call the graph at the bottom of Figure 6 an *uncrosser*. In any path 2-coloring of the uncrosser, the vertices labeled x are given the same color, as are the vertices labeled y. Furthermore, these two colors may be either the same or different.

Given these gadgets, we can present our reduction. By a *literal*, we mean any of $x_1, x_2, \ldots, x_m$ and their negations $\overline{x_1}, \overline{x_2}, \ldots, \overline{x_m}$. Let $C = \{c_1, c_2, \ldots, c_n\}$ be a set of clauses, where a *clause* is the logical-or of exactly three literals. Thus, C is an instance of 3-SAT. We shall construct in polynomial time a planar graph G_C with maximum degree 5 having the property that C is satisfiable if and only if G_C can be path 2-colored.

We begin constructing G_C with the vertices $x_1, x_2, \ldots, x_m, \overline{x_1}, \overline{x_2}, \ldots, \overline{x_m}$. Add a vertex labeled t. Now place between each x_i and $\overline{x_i}$ a negator. For each clause $x \vee y \vee z$ in C build a conjugator where the vertices x', y' and z' are attached by extenders to x, y, and z respectively. Furthermore, attach t to T and w by extenders. The resulting graph can be constructed in polynomial time. This graph might be non-planar, and some vertices may have large degree.

The graph we have constructed consists of one vertex corresponding to each literal, with pairs of these joined by negators, a vertex t, and a number of conjugators. The conjugators are joined to the rest of the graph via extenders;

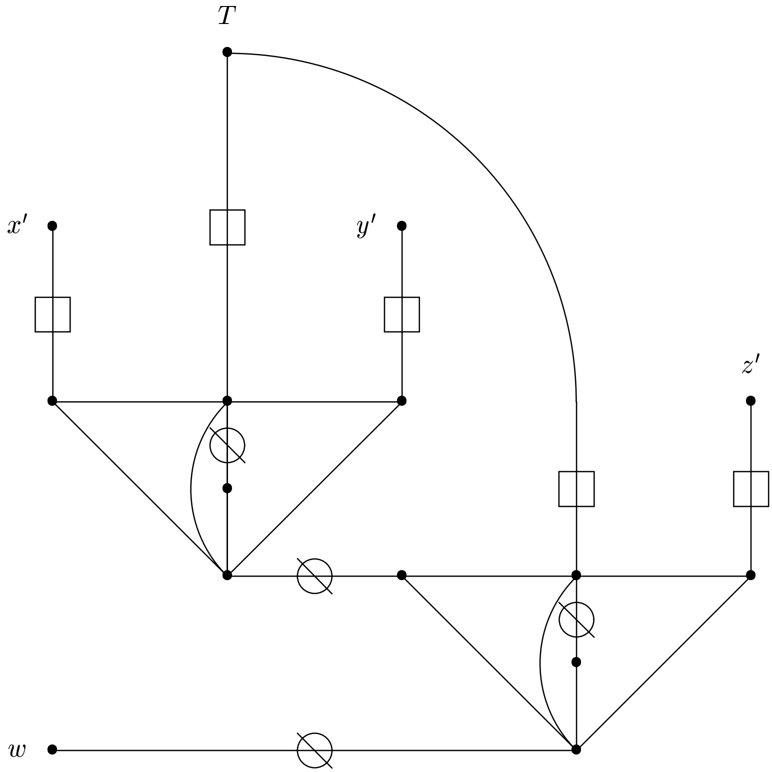


Fig. 5. The "conjugator", used in the proof of Theorem 2.8. If we identify the color assigned to T with **true** and the other color with **false**, then path 2-coloring this graph forces w to be colored with the logical-or of the colors assigned to x', y', and z'

if these extenders are removed, the resulting graph is planar. Thus, if our graph, as described above, is not planar, then it can be drawn in the plane so that only extenders cross. But crossed extenders can be replaced with an uncrosser, as illustrated in Figure 6, resulting in a planar graph. The total number of uncrossers used in this construction is $O(n^2)$ (recall that n is the number of clauses), and uncrossers can be built in polynomial time.

The vertices corresponding to literals, as well as the vertex t, may have high degree, due to a large number of extenders joining them with various conjugators. For each such vertex having degree greater than 5, replace the vertex with a "path" made of extenders, which we call a *splitter*, as shown in Figure 7. The resulting planar graph of maximum degree 5 is G_C. The total number of splitters used in this construction is $O(n)$, and splitters can be built in polynomial time. Hence, G_C can be built in polynomial time.

Let us now attempt to path 2-color this graph. If it is possible, identify the color assigned to t with **true**, and identify the other color with **false**.

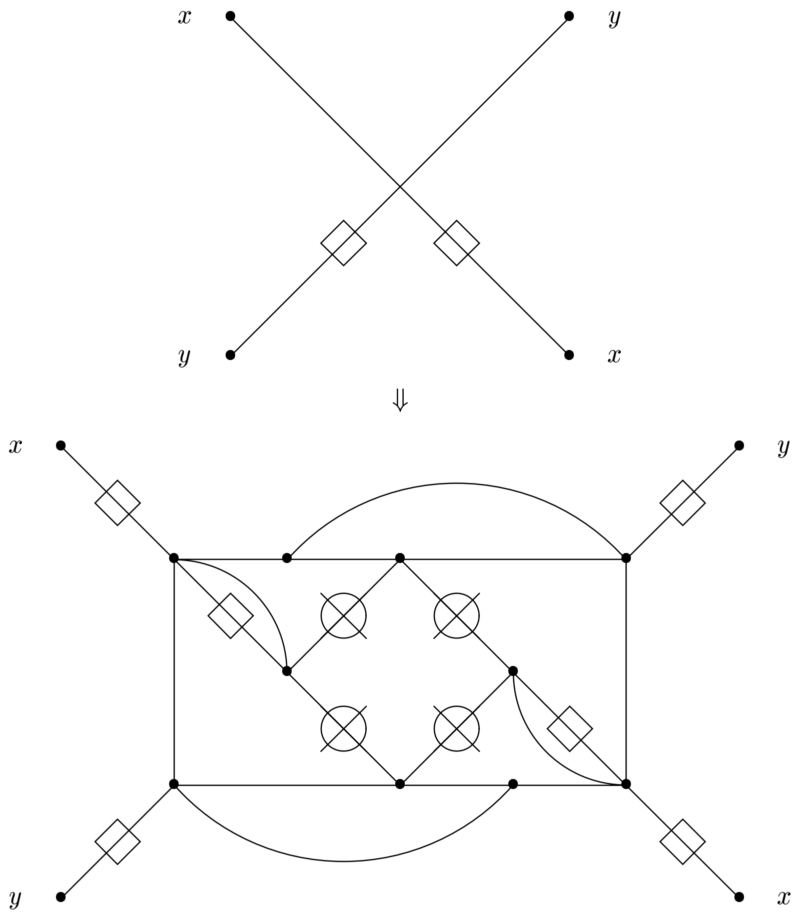


Fig. 6. The "uncrosser", used in the proof of Theorem 2.8. In any path 2-coloring of this graph, the vertices labeled x are given the same color, as are the vertices labeled y. Furthermore, these two colors may be either the same or different

Vertex w in a conjugator built for the clause $x \vee y \vee z$ will be given the label corresponding to the truth value of the clause. As w is joined by an extender to t, a truth assignment exists if and only if G_C has a path 2-coloring. □

It is possible that the above NP-completeness result continues to hold for triangle-free graphs. However, we have been unable to prove this. In fact we do not know whether there exists a triangle-free planar graph of maximum degree 5 whose path chromatic number is 3. However, we can show NP-completeness for triangle-free graphs if we raise the maximum degree to 6.

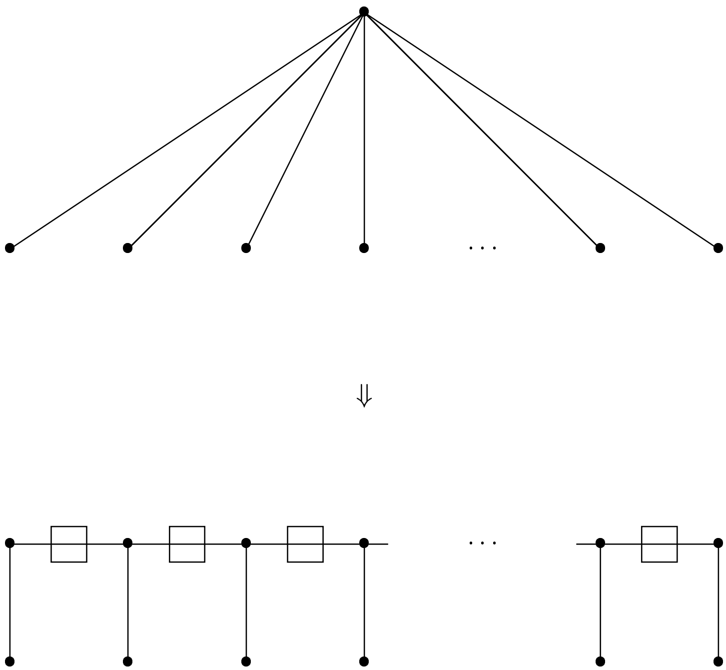


Fig. 7. Replacing a vertex of high degree with a *splitter*, that is, a "path" made of extenders. This technique is used in the proof of Theorem 2.8, to ensure a graph with maximum degree at most 5

Theorem 2.9. *It is NP-complete to determine whether a triangle-free planar graph with maximum degree* 6 *is path* 2*-colorable.*

Proof. Consider the graph at the top of Figure 8. In any path 2-coloring, vertices u and v must receive the same color. Let us use the design at the bottom of Figure 8 to represent this graph. Let H be the graph represented in Figure 9. This graph is planar with girth 4 and maximum degree 6. Furthermore, in any path 2-coloring, z and w must be given different colors.

Now, suppose G is a planar triangle-free graph with maximum degree at most 4. Form G' in the following manner. For each vertex v of G construct a copy of the graph in Figure 9 and make v adjacent to z and w. Graph G' can be path 2-colored if and only if G has a 1-defective 2-coloring. Furthermore, G' has maximum degree at most 6 and is planar and triangle-free.

By [GH03] the problem of deciding whether a planar triangle-free graph of maximum degree 4 has a 1-defective 2-coloring, is NP-complete. Hence, our desired result. □

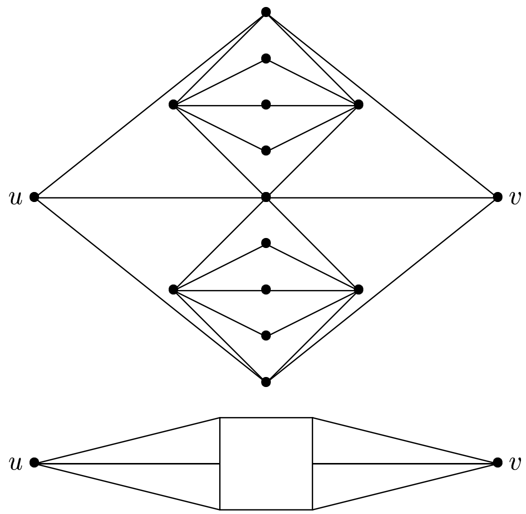


Fig. 8. A planar, triangle-free graph with maximum degree 6, along with a symbol representing it. Path 2-coloring this graph forces vertices u and v to receive the same color. This is used in the proof of Theorem 2.9

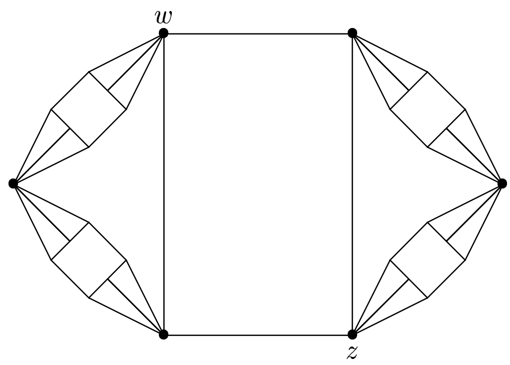


Fig. 9. A planar, triangle-free graph with maximum degree 6; path 2-coloring this graph forces vertices z and w to receive different colors. This is used in the proof of Theorem 2.9

3 Girth

Now we consider how the girth of a planar graph affects path 2-colorability and choosability.

Theorem 3.1. *If G is a planar graph with girth at least 6, then G is path 2-choosable. Furthermore, there is a polynomial-time algorithm to find the coloring.*

Proof. Let G be a planar graph with girth at least 6. We embed G in the plane. Assume that each vertex of G is assigned a list of 2 colors. We show that G can be path colored from these lists. We proceed by induction on the order of G.

We first deal with some easy cases. If G has either a vertex of degree 1 or less, or else two adjacent vertices of degree 2, then we may color every vertex other than these by the induction hypothesis. We then color each uncolored vertex with a color in its list that is not used on its already colored neighbor (if any). This is the required coloring.

We may thus assume that the minimum degree of G is at least 2, and no two vertices of degree 2 are adjacent.

The main portion of our proof uses the "vertex discharging method"; it will generally run as follows. We place a certain numerical value ("charge") on each vertex of G. We show that the sum of the vertex charges is positive. We then move the vertex charge around ("discharging"), without changing the sum of the charges. We use the existence of a vertex with positive charge to describe a certain structure in G. We remove this structure from G, color what remains using the induction hypothesis, and then color the vertices in the structure. Lastly, we alter this coloring to produce the required coloring of G.

Initial vertex charges. Let each vertex v of G be given a "charge" of $6 - 2d(v)$, where $d(v)$ is the degree of v.

We claim that the sum of these vertex charges is positive. To see this, let each face f of G be given charge $6 - \ell(f)$, where $\ell(f)$ is the length of f, and consider the sum of all vertex and face charges.

$$\sum_v [6 - 2d(v)] + \sum_f [6 - \ell(f)] = \sum_v 6 + \sum_f 6 - \left[\sum_v 2d(v) + \sum_f \ell(f)\right].$$

We are adding up 6 for each vertex of G and 6 for each face of G. For each edge of G, we then subtract 2 for each of its endpoints and 1 for each face with which it is incident (counting multiplicities), for a total of 6. If G has V vertices, F faces, and E edges, then the sum of all the charges is thus

$$6V + 6F - 6E = 6(V + F - E) = 12,$$

by Euler's Formula.

Since G has girth at least 6, each face has length at least 6, and therefore no face has positive charge. We conclude that the sum of all vertex charges is positive, as claimed.

Discharging. Since the sum of the vertex charges is positive, there must be a vertex with positive charge. Such vertices are precisely those having degree at most 2. Since we assumed the minimum degree of G is at least 2, we are left with vertices having degree exactly 2.

Now we move the charge around. We will always move a single unit of charge along an edge, subtracting 1 from the charge of one endpoint, and adding 1 to the charge of the other endpoint. When we do this, we will orient the edge in the direction the charge moved; we will never move charge along that edge again, nor will we alter the orientation. Thus, an edge is oriented if and only if a single unit of charge has been moved from its tail to its head. Furthermore, the total vertex charge remains constant.

Each vertex of degree 2 has a charge of 2. For each such vertex v, set the charge of v to zero and add 1 to the charge of each of its neighbors. Orient each edge incident with v away from v. Since there are no pairs of adjacent vertices of degree 2, this operation is well defined, and after it is performed, no vertex of degree 2 will ever have nonzero charge.

Next, we repeat the following charge redistribution operation, until one of the stopping conditions listed below is met. We find a vertex x of positive charge. We stop if

- vertex x has degree 3, or
- all edges incident with x are oriented toward it.

If we do not stop, then we claim that x has degree 4. It cannot have degree 2, since all such vertices now have charge zero. Nor can it have degree 3, or we would have stopped. It cannot have degree 5 since the only way to give such a vertex positive charge would be to add one unit of charge from each of its neighbors, in which case all incident edges would be oriented toward the vertex, and we would have stopped. Lastly, x cannot have degree 6 or more, since even adding one unit of charge to such a vertex from each of its neighbors would not be sufficient to give it positive charge.

Consider the number of incident edges that are oriented toward x. We claim there must be exactly 3 of these: if there were 4, then we would have stopped; if there were 2 or less, then x would not have positive charge. The remaining incident edge must be unoriented, or else x would not have positive charge. Thus, x has one incident unoriented edge and charge 1. We redistribute the charge by setting the charge of x to zero, increasing the charge of its neighbor along the unoriented edge by 1, and orienting this edge away from x.

As stated above, we repeat this process until one of the stopping conditions is met.

The above process must eventually end, since at each step we orient an unoriented edge. When the process ends, it is because one of the stopping conditions holds for a vertex with positive charge. We denote this vertex by x_0. We define a set S of vertices: for each vertex v of G, v lies in S if and only if there is a directed path from v to x, where we require a directed path to use no unoriented edges. Note that $x_0 \in S$, and there is at least one other vertex in S.

After the above process is complete, if a vertex has an incident edge oriented away from it, then every edge incident with this vertex must be oriented. Since every vertex in S, except for x_0, has an incident edge oriented away from it, we conclude that every edge whose endpoints both lie in S has been given an orientation. Denote by H the directed subgraph of G induced by S. Then H is connected, directed, and acyclic (as a digraph). The sources are vertices of degree 2 in G, which have out-degree either 1 or 2 in H. There is one sink, x_0, which has degree 3, 4, or 5 (since, as argued earlier, a vertex of degree 6 or more could never attain positive charge). All other vertices have degree 4 in G; in H they have in-degree 3 and out-degree 1.

Coloring. We now color G. We begin by removing S from $V(G)$, with one possible exception: if x_0 has degree 3 in G (that is, if we stopped due to the first stopping condition), then we do not remove x_0. Since we have removed at least one vertex from G, by the induction hypothesis we may path color the remaining vertices so that each vertex receives a color from its list.

Next, we place all the removed vertices (and edges) back into G. We shall color these vertices using an iterative procedure, to be described shortly. When this procedure is finished, we may not have a path coloring; if we do not, then we will fix the coloring. Below, an *in-neighbor* of a vertex v is a vertex adjacent to v along an edge oriented toward v.

First, we color each source in S with a color in its list that is not used by any neighbor in $G - S$. Since there is at most one such neighbor, the required color must exist.

Next, we iteratively find and color an uncolored vertex whose in-neighbors have all been colored (since H is a directed acyclic graph, such a vertex must exist, unless all vertices have been colored). If this vertex is not x_0, then it has degree 4 in G, and it has in-degree 3 and out-degree 1 in H. Since there are 2 colors in its list, there must be a color in the list that is used on at most one in-neighbor; we color the vertex with this color. On the other hand, if this vertex is x_0, then it has degree at most 5, and so there must be a color in its list that is used on at most 2 in-neighbors; we color x_0 with this color.

Fixing the coloring. When all vertices have been colored, we may or may not have a path coloring of G. There are two ways in which a vertex coloring may fail to be a path coloring: the existence of a monochromatic cycle, and the existence of a vertex with degree 3 or more in its color class. We now look at each of these in turn, and fix each problem if it occurs.

Consider monochromatic cycles. If an edge of G has one endpoint $a \in S$ and the other endpoint $b \notin S$, then either a has degree 2, in which case a and b have different colors, or else a is x_0 and it has degree 3. In the latter case, x_0 was colored along with the vertices not in S. We conclude that there can be no monochromatic cycle using some vertices in S and some not in S. Furthermore, there can be no monochromatic cycle using only vertices outside of S, by the induction hypothesis. We may, however, have a monochromatic cycle using only vertices in S. Due to our coloring procedure, any (undirected) cycle must contain a vertex that is a source in S. This must be a vertex of degree 2 in G that is given the same color as both of its neighbors. We fix this by switching its color to the other color in its list, so that it now has degree zero in its color class. In doing this, we create no new cycles; nor do we increase the degree of any vertex in its color class. We may thus repeat until all monochromatic cycles have been eliminated.

Now consider vertices with high degree in their color class. Again, when we path colored using the induction hypothesis, we did not create any such vertices. Furthermore, our coloring method did not increase the degree, in its color class, of any vertex not in S. There may, however, be a vertex in S that has more than 2 neighbors in its color class. Such a vertex cannot be a source, since these have degree 2 in G. Nor can it be one of the vertices with in-degree 3 and out-degree 1, since these were colored so as to have the same color as at most one of their in-neighbors; thus, such a vertex has degree at most 2 in its color class. Nor can such a vertex be x_0, if the second stopping condition was used, since, in this case, all neighbors are in-neighbors, and x_0 was colored so as to have the same color as at most two of its in-neighbors.

Thus, the only vertex that can have degree greater than 2 in its color class is x_0, and this can only happen if the first stopping condition was used, that is, if x_0 has degree 3. We fix this by switching the color of x_0 to the other color in its list, so that it now has degree zero in its color class. In doing this, we create no new cycles; nor do we increase the degree of any vertex in its color class.

Thus, we have the required coloring.

Algorithmic efficiency. We now consider the above coloring method as an algorithm, and show that it is polynomial-time.

Say the original graph G has n vertices and m edges. Since G is planar, m is $O(n)$. Each call to the algorithm first searches for either a vertex of degree at most 1, or two adjacent vertices of degree 2. This requires $O(n)$ steps. If one of these is found, then a recursive call is made, and, after a constant number of operations, the algorithm ends.

Otherwise, the algorithm places charge on all the vertices, requiring $O(n)$ steps. Vertices of degree 2 are discharged: $O(n)$ steps. Then the algorithm repeatedly searches for a vertex with positive charge. If this vertex does not satisfy a stopping condition, then the charge is rearranged. The requires $O(n)$ steps to be done once; it is done at most $O(n)$ times, resulting in $O(n^2)$ steps total. Next, there is one recursive call. After this, the coloring is finished and then fixed, requiring $O(n)$ steps. Thus, aside from the recursive call, the algorithm requires $O(n^2)$ steps.

Each call to the algorithm makes at most one recursive call, and each recursive call removes at least one vertex. The recursion depth is thus $O(n)$, and so the order of the algorithm is $O(n^3)$, which is polynomial-time. ◻

We can prove similar results for graphs on surfaces of higher genus. In order to do this, we define a list-coloring version of 1-defective coloring, just as we did for path coloring. A graph G is 1-*defective* 2-*choosable* if, whenever we assign lists of 2 colors to each vertex of G, there is a 1-defective coloring of G in which each vertex receives a color from its list. We note that, if a graph is 1-defective 2-choosable, then it is path 2-choosable.

We will make use of the following result of Galluccio, Goddyn, and Hell [GGH01, Thm. 3.2].

Theorem 3.2 (Galluccio, Goddyn, Hell 2001 [GGH01]). *Let g be a nonnegative integer, and let G be a graph that embeds on an orientable surface of genus g. If G has minimum degree at least 3, then the girth of G is at most*

$$4 + \lfloor 2\log_2(g + 3/2) \rfloor .$$

◻

We denote the value of the bound in Theorem 3.2 by k_g.

Theorem 3.3. *Let g be a nonnegative integer, and let G be a graph that embeds on an orientable surface S of genus g. If G has girth at least $2k_g + 1 = O(\log g)$, then G is 1-defective 2-choosable, and therefore is path 2-choosable. Furthermore, there is a polynomial-time algorithm to find the coloring.*

Proof. Fix g and S, and let G be a graph that embeds on S and has girth at least $2k_g + 1$. Assume that each vertex of G is assigned a list of 2 colors. We show that G has a 1-defective coloring in which each vertex receives a color from its list. We proceed by induction on the order of G.

If G is not connected, then we apply the induction hypothesis to each component of G, and we are done.

We claim that G has either a vertex of degree 1 or less, or else two adjacent vertices of degree 2. If this is the case, then we may color every vertex other than these by the induction hypothesis. We then color each uncolored vertex with a color in its list that is not used on its already colored neighbor (if any). This is the required coloring.

To verify the claim, suppose that G has minimum degree at least 2. We need to show that G has two adjacent vertices of degree 2. To see this, create a possibly non-simple graph G' based on G by iteratively removing a vertex of degree 2 in G and replacing it with an edge between its two neighbors, until no vertices of degree 2 remain. This graph G' cannot be a forest, since G has minimum degree at least 2. Therefore, let C be a shortest cycle of G', and let ℓ be the length of C. If G' is simple, then, since G has minimum degree at least 3 and embeds on S, we may apply Theorem 3.2 to obtain

$$\ell \leq k_g.$$

If G' is not simple, then either $\ell = 1 \leq k_g$ (a loop), or $\ell = 2 \leq k_g$ (a pair of parallel edges). Thus, the above bound on ℓ holds for both simple and non-simple G'. Since the girth of G is at least $2k_g + 1$, we conclude that the girth of G is at least $2\ell + 1$. Cycle C must therefore result from the removal of at least $\ell + 1$ vertices of degree 2. By the Pigeonhole Principle, at least two of those removed vertices must have been removed from the same edge of C, and so the claim is proven.

It remains to verify that there exists a polynomial-time algorithm to produce this coloring. We can create a recursive algorithm by following the steps in the above inductive proof.

Say the original graph G has n vertices. Each call to the algorithm first finds either a vertex with degree at most 1, or two adjacent vertices of degree 2. This may require looking at $O(n)$ vertices. Then a single recursive call is made. After the recursive call, the coloring is completed, using a constant number of steps. Thus, aside from the recursive call, the algorithm requires $O(n)$ steps. The recursion depth is $O(n)$, and so the order of the algorithm is $O(n^2)$, which is polynomial-time. ◻

We have shown that graphs of high girth, on the plane and other surfaces, are path 2-choosable. On the other hand, as we shall see, planar graphs of low girth may not be path 2-colorable. Furthermore, given a planar graph of low girth, it is hard to determine whether the graph is path 2-colorable. The following corollary follows immediately from Theorem 2.9.

Corollary 3.4. *It is NP-complete to determine whether a planar graph with girth* 4 *is path* 2*-colorable.* ◻

We note that path 2-colorability of planar graphs with girth 5 remains open.

Problem 3.5. Let G be a planar graph of girth 5. Is G necessarily path 2-colorable? Path 2-choosable? What about algorithmic issues? ◻

We close with an example of a planar graph that has no path 2-coloring, and, in a sense, "almost" has girth 5. This graph is triangle free and $K_{2,3}$-free, that is, it has no subgraph isomorphic to $K_{2,3}$. We note that a graph has girth at least 5 if and only if it is triangle-free and $K_{2,2}$-free.

Example 3.6. There exists a planar graph of girth 4 that has no path 2-coloring, and has no subgraph isomorphic to $K_{2,3}$.

Let B be the graph shown in Figure 10. We use B to construct a planar graph G of order 169 that meets the above conditions.

We construct G as follows. Begin with three copies of B. For each vertex v on the outer face of each copy, add two new vertices and join them to v. Lastly,

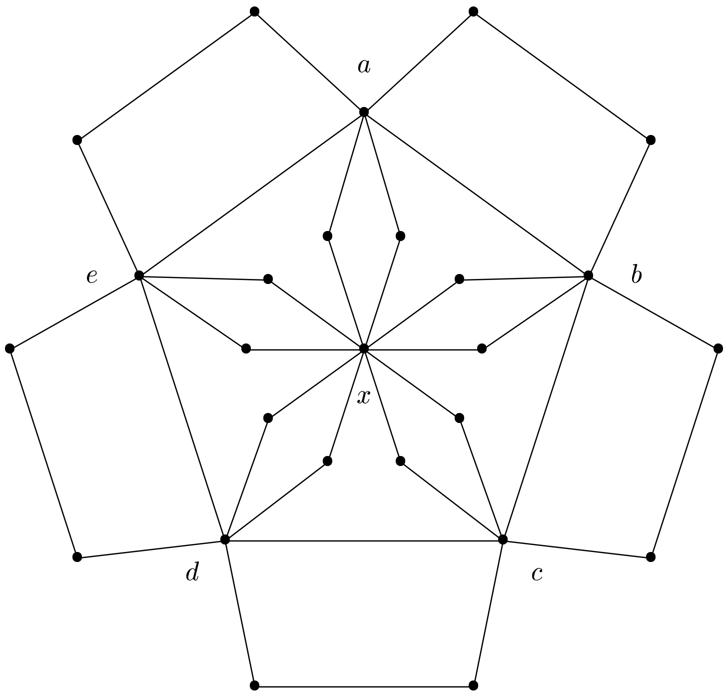


Fig. 10. Graph "B", used in Example 3.6. There is no path 2-coloring of B with colors red and blue, in which each red vertex on the outer face has no red neighbors in B. Graph B is used to construct a planar graph of order 169 that has girth 4, is $K_{2,3}$-free, and has no path 2-coloring

add a vertex z and join it to each of the new vertices. Let G be the resulting graph. Clearly, G has girth 4 and no subgraph isomorphic to $K_{2,3}$.

We now prove that G has no path 2-coloring. We claim (and we will show below) that B has no path 2-coloring, using colors red and blue, in which each red vertex on the outer face has no red neighbors in B. Now, suppose we attempt to path 2-color G. We may assume vertex z is colored blue. At most two neighbors of z are colored blue. Thus, for at least one of the three copies of B all of the attached pendants ("new vertices" above) are red. This means that each red vertex in the outer face of this copy of B already has two red neighbors; it can have no other red neighbors in this copy of B. However, by the above claimed property, B cannot be path 2-colored in this manner. Thus, G has no path 2-coloring.

We now verify the claimed property of graph B.

Suppose we attempt to path 2-color B, using colors red and blue, so that each red vertex on the outer face has no red neighbors. Consider the 5-cycle with vertices a, b, c, d, e. This cycle cannot be completely blue. Nor can it contain more than two red vertices, since we cannot have adjacent red vertices on the outer face. We thus have two cases: cycle $abcde$ has exactly one red vertex, or it has exactly two red vertices, which are not adjacent.

In the first case, without loss of generality suppose vertex a is red, and b, c, d, e are all blue. Then c and d each have two blue neighbors; all their other neighbors must be red. But vertices c and d lie in a 4-cycle, all of whose vertices lie on the outer face. The other two vertices in this cycle must be red, resulting in adjacent red vertices on the outer face, which is a contradiction.

In the second case, without loss of generality suppose vertices b and e are red, with a, c, d being blue. Since b and e lie on the outer face, they can have no red neighbors. Each of b, e has two neighbors that are not on the outer face; these neighbors must be blue. Vertex x thus has at least four blue neighbors; x must be red, and so it can have at most two red neighbors.

Consider the neighbors of x that are also neighbors of c or d. At most two of these can be colored red, since x may have at most two red neighbors. However, vertices c and d are both colored blue, and both have a blue neighbor (each other). Thus, c and d may each have at most one other blue neighbor. We see that x must have exactly two red neighbors, one adjacent to c, and the other adjacent to d. The other common neighbor of x and c, and the other common neighbor of x and d, must be blue.

Thus, c and d each have two blue neighbors. As in the first case, vertices c and d lie in a 4-cycle, all of whose vertices lie on the outer face. The other two vertices in this cycle must be red, resulting in adjacent red vertices on the outer face, which is a contradiction.

We have verified the claimed property of B, which completes the argument. □

References

[AEGW89] J. Akiyama, H. Era, S. V. Gervacio, and M. Watanabe, Path chromatic numbers of graphs, *J. Graph Theory* **13** (5) (1989), 569–575.

[AJ85] J. Andrews, M. Jacobson, On a generalization of chromatic number, *Congressus Numer.* **47** (1985), 33–48.

[AH77] K. Appel and W. Haken, Every planar map is four colorable. Part I. Discharging, *Illinois J. Math.* **21** (1977), 429–490.

[AHK77] K. Appel, W. Haken, and J. Koch, Every planar map is four colorable. Part II. Reducibility, *Illinois J. Math.* **21** (1977), 491–567.

[BKT00] O. V. Borodin, A. V. Kostochka, B. Toft, Variable degeneracy: extensions of Brooks' and Gallai's theorems, *Discrete Math.* **214** (2000), no. 1-3, 101–112.

[CL96] G. Chartrand, L. Lesniak, *Graphs & Digraphs*, 3rd ed., Chapman & Hall, London, 1996.

[CGJ97] L. Cowen, W. Goddard, C. E. Jesurum, Defective coloring revisited, *J. Graph Theory* **24** (1997), 205–219.

[CCW86] L. Cowen, R. H. Cowen, D. R. Woodall, Defective colorings of graphs in surfaces: partitions into subgraphs of bounded valency, *J. Graph Theory* **10** (1986), 187–195.

[Cow93] R. Cowen, Some connections between set theory and computer science, in *Computational logic and proof theory* (Brno, 1993), Lecture Notes in Comput. Sci., 713, Springer, Berlin, 1993, pp. 14–22.

[Far04] A. Farrugia, Vertex-partitioning into fixed additive induced-hereditary properties is NP-hard, *Elec. J. Combin.* **11** (2004), no. 1, research paper 46, 9 pp.

[GGH01] A. Galluccio, L. A. Goddyn, P. Hell, High-girth graphs avoiding a minor are nearly bipartite. *J. Combin. Theory Ser. B* **83** (2001), no. 1, 1–14.

[GH03] J. Gimbel, C. Hartman, Subcolorings and the subchromatic number of a graph, *Discrete Math.* **272** (2003), no. 2-3, 139–154.

[God91] W. Goddard, Acyclic colorings of planar graphs, *Discrete Math.* **91** (1991), 91–94.

[Har97] C. Hartman, *Extremal Problems in Graph Theory*, Ph.D. Thesis, University of Illinois, 1997.

[Poh90] K. S. Poh, On the linear vertex-arboricity of a planar graph, *J. Graph Theory* **14** (1990), 73–75.

[RSST97] N. Robertson, D. P. Sanders, P. D. Seymour, and R. Thomas, The four colour theorem, *J. Combin. Theory Ser. B.* **70** (1997), 2–44.

Chromatic Numbers and Homomorphisms of Large Girth Hypergraphs

Dwight Duffus[1], Vojtěch Rödl[1], Bill Sands[2], and Norbert Sauer[2]

[1] Emory University, Mathematics & Computer Science Department, Atlanta GA 30322, USA
dwight@mathcs.emory.edu, rodl@mathcs.emory.edu
[2] University of Calgary, Mathematics & Statistics Department, 2500 University Dr. NW., Calgary AB T2N1N4, Canada
sands@math.ucalgary.ca, nsauer@math.ucalgary.ca

Summary. We consider the problem of determining the minimum chromatic number of graphs and hypergraphs of large girth which cannot be mapped under a homomorphism to a specified graph or hypergraph. More generally, we are interested in large girth hypergraphs that do not admit a vertex partition of specified size such that the subhypergraphs induced by the partition blocks have a homomorphism to a given hypergraph. In the process, a general probabilistic construction of large girth hypergraphs is obtained, and general definitions of chromatic number and homomorphisms are considered.

AMS Subject Classification. 05C15, 05C65.
Keywords. Hypergraph colouring, hypergraph homomorphisms, vertex partitions.

1 Introduction

This paper is motivated by a problem of N. Sauer and R. Winkler. Given a graph G $= (V, E)$, find the smallest chromatic number of a triangle-free graph H so that after removing any induced subgraph H$[V']$ which has a homomorphism to G the remaining graph H$[V - V']$ still does not have a homomorphism to G. We provide an answer to this [see Theorems 1.3, 1.4 and 5.4] and in the process are led to two observations. First, our investigations are related to a recent paper by J. Nešetřil and X. Zhu [NZ04] that connects the well-known results of P. Erdős [Erd59] on the existence of graphs with large girth and high chromatic number to the study of homomorphisms. Second, our methods both use hypergraph constructions and yield results for hypergraph versions of the problem. This leads to consideration of how best to define chromatic numbers and homomorphisms of hypergraphs.

Graph homomorphisms have emerged as a fruitful tool within graph theory. In fact, there is a new book by Nešetřil and P. Hell [HN04] devoted to this subject that highlights diverse applications. However, homomorphisms of hypergraphs, and their relationship to possible definitions of chromatic number, have not been studied extensively. One of the objectives of this work is to examine these notions in the hypergraph setting.

Let us start off by stating the Nešetřil–Zhu result, showing how it is related to the Sauer–Winkler problem, and seeing that the most familiar or obvious ways to define chromatic number and homomorphisms for hypergraphs lead to very different results for the graph and hypergraph cases.

We first consider a simple version of the question of Sauer and Winkler. Given a graph A, what is the minimum chromatic number $\psi(\mathrm{A})$ of a graph G such that there is no homomorphism from G to A? It is easy to see that $\psi(\mathrm{A}) = \omega(\mathrm{A}) + 1$, where $\omega(\mathrm{A}) = \omega$ is the clique number of A. The complete graph on $\omega + 1$ vertices, $\mathrm{K}_{\omega+1}$, has no homomorphism into A, since the image of a clique under a graph homomorphism is a clique of the same size. On the other hand, any graph G of chromatic number ω has a homomorphism, defined by an ω-colouring of G, into a clique of A. This easy observation can be strengthened considerably with the Nešetřil–Zhu theorem, alluded to above.

Theorem 1.1 (J. Nešetřil and X. Zhu, [NZ04]). *For every graph* H *and for all positive integers n and l there exists a graph* G *with the following properties:*

1. girth(G) $> l$.
2. *For every graph* H' *with at most n vertices, there exists a homomorphism* $g : \mathrm{G} \to \mathrm{H}'$ *if and only if there exists a homomorphism* $f : \mathrm{H} \to \mathrm{H}'$.

In Section 5 we show how this result gives the following.

Theorem 1.2. *Let* A *be a graph and let l be an integer. Then there is a graph* G *with* girth(G) $> l$ *and* $\chi(\mathrm{G}) = \psi(\mathrm{A})$ *which does not have a homomorphism to* A.

Now, let us generalize ψ to capture the Sauer–Winkler problem. Say that a graph G is *s-partition homomorphic* to a graph A if there exists a partition $V(\mathrm{G}) = V_1 \cup V_2 \cup \ldots \cup V_s$ such that for all i there is a homomorphism of $\mathrm{G}[V_i]$ to A. Let $\psi(\mathrm{A}, s)$ be the least chromatic number of a graph which is not s-partition homomorphic to A. Then $\psi(\mathrm{A}, 1) = \psi(\mathrm{A})$ and $\psi(\mathrm{A}, 2)$ is the least chromatic number of a graph $\mathrm{G} = (V, E)$ such that whenever an induced subgraph $\mathrm{G}[V']$ has a homomorphism to A, then $\mathrm{G}[V - V']$ does not — the original problem.

Without a girth restriction, just about the same reasoning as in the case $s = 1$ shows that $\psi(\mathrm{A}, s) = s \cdot \omega(\mathrm{A}) + 1$. But we are also able to extend the result, just as in Theorem 1.2.

Theorem 1.3. *Let* A *be a graph and let l, s be integers. Then there is a graph* G *with* girth(G) $> l$ *and* $\chi(\mathrm{G}) = \psi(\mathrm{A}, s)$ *which is not s-homomorphic to* A.

One way to prove this is based on a hypergraph packing construction, a probabilistic construction of a large girth hypergraph [with additional properties], and Theorem 1.1. This argument is outlined in Section 5. In the process, we obtained an extension of Theorem 1.3 from the graph case to the k-uniform hypergraph case. Before stating this, here is some language needed for hypergraphs.

A *hypergraph* H is a pair (V, E) where E is a family of subsets of the *vertex set* V. To avoid trivialities, assume that the members of E, that is, the *edges* of H all have size at least 2. Call a hypergraph *simple* if edges intersect in at most one vertex. Call $\mathrm{H} = (V, E)$ a *k-uniform* hypergraph if E is a family of k-element subsets of V; thus, a usual (simple, undirected) graph is a 2-uniform hypergraph.

Given a hypergraph H, we often let $V(\mathrm{H})$ denote the set of vertices of H and $E(\mathrm{H})$ stand for the set of edges of H. Given $S \subseteq V(\mathrm{H})$, $\mathrm{H}[S]$ is the hypergraph with vertex set S and edges $\mathfrak{P}(S) \cap E(\mathrm{H})$, where $\mathfrak{P}(S)$ is the power set of S. Call $\mathrm{H}[S]$ the *subhypergraph of* H *induced by* S.

An *r-colouring* of H is a map with domain $V(\mathrm{H})$, range within an r-element set, and which is not constant on any edge of H. The *(weak) chromatic number* of H, denoted by $\chi(\mathrm{H})$, is the least r for which H has an r-colouring. Note that $\chi(\mathrm{H})$ exists since every edge has at least two elements.

Given k-uniform hypergraphs F_1 and F_2, a mapping $f : V(\mathrm{F}_1) \mapsto V(\mathrm{F}_2)$ is a *homomorphism* of F_1 to F_2 if $\{f(x) : x \in e\} \in E(\mathrm{F}_2)$ for every $e \in E(\mathrm{F}_1)$. For example, if G is a graph, then $\chi(\mathrm{G}) \leq r$ if and only if the graph G has a homomorphism to K_r.

We will be concerned with hypergraphs of large girth. A *circuit of length c* in a hypergraph $\mathrm{H} = (V, E)$ is a sequence $e_1, e_2, \ldots, e_c$ of distinct members of E and distinct elements $v_1, v_2, \ldots, v_c$ of V such that

$$v_i \in e_i \cap e_{i+1} \ (i = 1, 2, \ldots, c-1), \text{ and } v_c \in e_1 \cap e_c.$$

Thus, if H is k-uniform then each circuit of length c corresponds to a family of distinct k-sets whose union contains at most $c(k-1)$ elements of V. Conversely, it is well-known that given a family of c members of E with union of size at most $c(k-1)$ some subfamily constitutes a circuit of length at most c. [See, for instance, [EH66]].

The k-uniform hypergraph F is *s-partition homomorphic* to the k-uniform hypergraph A if there exists a partition $V(\mathrm{F}) = V_1 \cup V_2 \cup \ldots \cup V_s$ such that for all i there is a homomorphism of $\mathrm{F}[V_i]$ to A. Let $\psi(\mathrm{A}, s)$ be the least chromatic number of a k-uniform hypergraph which is not s-partition homomorphic to A.

Before stating the result for k-uniform hypergraphs, we note a connection of the function $\psi(\mathrm{A}, s)$ to a quite natural generalized chromatic number, which was suggested by one of the referees. Given k-uniform hypergraphs A and H, let $\chi_{\mathrm{A}}(\mathrm{H})$ be the minimum t for which there is a partition $V(\mathrm{H}) = V_1 \cup V_2 \cup \ldots \cup V_t$ such that for all i there is a homomorphism of $\mathrm{H}[V_i]$ to A. Of course, $\chi(\mathrm{H}) = \chi_{\mathrm{K}_1}(\mathrm{H})$. Also, it is easy to see that

$$\psi(\mathrm{A}, s) = \min\{\chi(\mathrm{H}) \ : \ \chi_{\mathrm{A}}(\mathrm{H}) \geq s+1\}.$$

Theorem 1.4. *Let* A *be a k-uniform hypergraph and l, s be positive integers. Then*

$$\psi(\mathrm{A}, s) = \begin{cases} s \cdot \omega(\mathrm{A}) + 1 & \textit{if } \mathrm{A} \textit{ is a graph,} \\ s+1 & \textit{if } k \geq 3. \end{cases}$$

Moreover, for all l there is a k-uniform hypergraph F *with* girth(F) $> l$ *and* $\chi(\mathrm{F}) = \psi(\mathrm{A}, s)$ *which is not s-partition homomorphic to* A.

The difference between the graph and hypergraph results is due to the fact that the usual definitions of both chromatic number and homomorphism are not broad enough in the hypergraph setting.

We present an approach in the following three sections that provides more general definitions of chromatic number and homomorphism, a main result that subsumes both preceding theorems [Theorem 2.1], a quite general probabilistic hypergraph construction [Lemma 2.2], and proofs of the theorem and lemma.

In Section 5, we present an alternative proof of Theorem 1.2 based on the Nešetřil–Zhu result. We also introduce the hypergraph packing construction and, with it, lift the results for $s = 1$ to the general setting and a proof of Theorem 1.4. The second version of Theorem 1.4, Theorem 5.4, is somewhat sharper. The reader can find the required terminology in Section 5, independent of the intervening sections.

In Section 6, we close with a constructive approach to some special cases.

2 New Definitions, the Main Theorem and the Probabilistic Lemma

In Theorem 1.4, we see the difference between the graph case and the k-uniform hypergraph case, for $k \geq 3$. In order to create a more general result that subsumes both cases, we amend the definitions of homomorphism and chromatic number to better suit the hypergraph setting.

In the sequel, all hypergraphs are finite and k-uniform, $k \geq 2$. Let $1 < p, q \leq k$ and let $\mathrm{H}_i = (V_i, E_i)$, $i = 0, 1$, be hypergraphs.

(1) A *p-homomorphism* f of H_0 to H_1 is a map of V_0 to V_1 such that for all $e_0 \in E_0$ there exists $e_1 \in E_1$ such that

$$f[e_0] \subseteq e_1, \text{ and } |f[e_0]| \geq p \ ;$$

write $\mathrm{H}_0 \xrightarrow{p} \mathrm{H}_1$ if there exists a p-homomorphism of H_0 to H_1, and $\mathrm{H}_0 \overset{p}{\nrightarrow} \mathrm{H}_1$ if not.

(2) $\chi^{(q)}(\mathrm{H}_0) = \min\{\chi \mid \text{there exists } c : V_0 \to [\chi], \text{ for all } e \in E_0, \ |c[e]| \geq q\}$

(3) $\psi_{p,q}(\mathrm{H}_1) = \min\{\chi^{(q)}(\mathrm{H}_0) \mid \mathrm{H}_0 \overset{p}{\nrightarrow} \mathrm{H}_1\}$

Our results hold in the more general case of s-partitions.

(4) Say that H_0 is *s-partition p-homomorphic* to H_1 if there is a partition $X_1 \cup X_2 \cup \ldots \cup X_s$ of the vertices of H_0 such that for all i, $\mathrm{H}_0[X_i] \xrightarrow{p} \mathrm{H}_1$.
(5) Let $\psi_{p,q,s}(\mathrm{H}_1)$ be the minimum $\chi^{(q)}(\mathrm{H}_0)$ such that H_0 is not s-partition p-homomorphic to H_1.

Here is the main result. For a k-uniform hypergraph $\mathrm{H} = (V, E)$, we let the *clique number* $\omega(\mathrm{H})$ be the maximum integer ω such that there is an ω-subset V' of V all of whose k-subsets of V' are members of E.

Theorem 2.1. *Let k, p, q, s and l be positive integers and let* A *be a k-uniform hypergraph.*

(1) *For $1 < p \leq q \leq k$, let $q = d(p-1)+r$, where $0 < r \leq p-1$, let $\omega = \omega(\mathrm{A})$ and set $t = d\omega + r - 1$. Then $\psi_{p,q,s}(\mathrm{A}) = st + 1$.*
(2) *For $1 < q < p \leq k$, $\psi_{p,q,s}(\mathrm{A}) = s(q-1) + 1$.*

Moreover, there is a k-uniform hypergraph H *with* $\mathrm{girth}(\mathrm{H}) > l$ *such that* $\chi^{(q)}(\mathrm{H}) = \psi_{p,q,s}(\mathrm{A})$ *and* H *not s-partition p-homomorphic to* A.

This theorem contains the results in the introduction. The usual definition of [weak] chromatic number has $q = 2$ and the definition of homomorphism used in Section 1 sets $p = k$. The graph case corresponds to $p = q = 2$, so Theorem 2.1 states that $\psi(\mathrm{A}, s) = st + 1 = s\omega + 1$, as given in Theorem 1.4. For $k \geq 3$, we have $q = 2 < 3 \leq k = p$, so the second part of the theorem yields $\psi(\mathrm{A}, s) = s + 1$, also as in Theorem 1.4.

The proof of Theorem 2.1 is given in Section 3. It depends upon a general probabilistic lemma which provides all the "constructions" we require. We state this now, but delay the proof until Section 4.

Call a sequence $\vec{k} = (k_1, k_2, \ldots, k_c)$ of positive integers a *k-sequence* if $\sum_{\alpha=1}^{c} k_\alpha = k$.

Lemma 2.2. *Given $\epsilon > 0$, positive integers m, k, l, h, and k-sequences*

$$\vec{k}^{(\alpha)} = (k_1^{(\alpha)}, k_2^{(\alpha)}, \ldots, k_{c_\alpha}^{(\alpha)}),\ \alpha = 1, 2, \ldots, h,$$

there exists n_0 such that for all $n \geq n_0$, there is a k-uniform hypergraph H *with vertex set*

$$V(\mathrm{H}) = \bigcup_{i=1}^{m+1} V_i,\ |V_i| = n\ (i = 1, 2, \ldots, m+1)$$

such that

(1) *for all $e \in E(\mathrm{H})$ there exist α and a sequence $1 \leq j_1 < j_2 < \ldots < j_{c_\alpha} \leq m+1$ such that for all $\beta = 1, 2, \ldots, c_\alpha$, $|e \cap V_{j_\beta}| = k_\beta^{(\alpha)}$;*
(2) *for all $\alpha = 1, 2, \ldots, h$, for all $1 \leq j_1 < j_2 < \ldots < j_{c_\alpha} \leq m+1$, for all $\beta = 1, 2, \ldots, c_\alpha$, and for all $V'_{j_\beta} \subseteq V_{j_\beta}$ with $|V'_{j_\beta}| \geq \epsilon n$, there exists $e \in E(\mathrm{H})$ such that $|e \cap V'_{j_\beta}| = k_\beta^{(\alpha)}$; and*
(3) $\mathrm{girth}(\mathrm{H}) > l$.

3 The Proof of the Main Theorem

Proof of Theorem 2.1 (1)

Let $\mathrm{A} = (V, E)$, and integers k, p, q, s, l and ω be as in the theorem statement, and, also as above, let $q = d(p-1)+r$, where $0 < r \le p-1$, and $t = d\omega + r - 1$.

We first prove that $\psi_{p,q,s}(\mathrm{A}) > st$ by showing if H is any k-uniform hypergraph with $\chi^{(q)}(\mathrm{H}) = st$ then there is a partition $V(\mathrm{H}) = Y_1 \cup Y_2 \cup \ldots \cup Y_s$ such that for all i, $\mathrm{H}^{(i)} := \mathrm{H}[Y_i] \xrightarrow{p} A$. To this end, let $c : V(\mathrm{H}) \to [st]$ be a q-colouring of H with classes $C_1, C_2, \ldots, C_{st}$. That is, for all $e \in E(\mathrm{H})$, $e \cap C_j \neq \emptyset$ for at least q distinct j's.

Group these st sets into s, say,

$$Y_i = C_{(i-1)t+1} \cup \ldots \cup C_{it}, \ (i = 1, 2, \ldots, s).$$

We show that this is the required s-partition of H.

Let $W = \{w_1, w_2, \ldots, w_\omega\} \subseteq V$ induce an ω-clique in A. We argue that there is a p-homomorphism f of $\mathrm{H}^{(i)}$ to A with range contained in W. Because the vertex set of $\mathrm{H}^{(i)}$ is the union of t classes C_j and $t = d\omega + r - 1$, we can group the vertices of $\mathrm{H}^{(i)}$ into disjoint subsets $D_1, D_2, \ldots, D_\omega$ as follows:

each of $D_1, \ldots, D_{r-1}$ is the union of $d+1$ distinct C_j's,
each of $D_r, \ldots, D_\omega$ is the union of d distinct C_j's,

and each C_j, $j = (i-1)t+1, \ldots, it$, is contained in one D_m. This is possible since $(r-1)(d+1) + (\omega - r + 1)d = d\omega + r - 1 = t$.

Define $f : V(\mathrm{H}^{(i)}) \to W$ by $f(v) = w_m$ for all $v \in D_m$, $m = 1, 2, \ldots, \omega$. Proving that f is a p-homomorphism requires that for all $e \in E(\mathrm{H}^{(i)})$, $f[e]$ is contained in an edge of A and $|f[e]| \ge p$. The former is true because $|f[e]| \le k$ and W induces a clique in A. If the latter fails then there are $m_1, \ldots, m_{p-1}$ such that

$$f[e] \subseteq \{w_{m_1}, \ldots, w_{m_{p-1}}\},$$

$$e \subseteq D_{m_1} \cup \ldots \cup D_{m_{p-1}}, \text{ and therefore}$$

$$C_{j_1} \cup \ldots \cup C_{j_q} \subseteq D_{m_1} \cup \ldots \cup D_{m_{p-1}},$$

where $e \cap C_{j_u} \neq \emptyset$ for $u = 1, \ldots, q$. This implies that at most $p-1$ D_m's contain at least q C_j's. This is impossible because $p-1$ of the D_m's contain at most

$$(r-1)(d+1) + (p-r)d = (p-1)d + r - 1 = q - 1$$

distinct C_j's. Thus f is a p-homomorphism.

In order to prove that $\psi_{p,q,s}(\mathrm{A}) \le st+1$, we use Lemma 2.2 to provide an example of a k-uniform hypergraph $\mathrm{H}_{st+1} = (V(\mathrm{H}_{st+1}), E(\mathrm{H}_{st+1}))$ such that $\chi^{(q)}(\mathrm{H}_{st+1}) = st+1$ and for all partitions $V(\mathrm{H}_{st+1}) = X_1 \cup \ldots \cup X_s$ there is some i such that $\mathrm{H}^{(i)} := \mathrm{H}_{st+1}[X_i]$ is not p-homomorphic to A. Moreover, this hypergraph also has girth greater than any specified l. To use Lemma 2.2, we also have to specify parameters m and h, the value of ϵ and the k-sequences $\vec{k}^{(\alpha)} = (k_1^{(\alpha)}, k_2^{(\alpha)}, \ldots, k_{c_\alpha}^{(\alpha)})$, $\alpha = 1, 2, \ldots, h$.

Let $m = st$, $h = 2$, $\epsilon = 1/(st|V|)$ and specify the two k-sequences to be

$$\vec{k}^{(1)} = (1, 1, \ldots, 1), \text{ and } \vec{k}^{(2)} = (k - q + 1, 1, \ldots, 1).$$

With the notation of Lemma 2.2, $c_1 = k$ and $c_2 = q$.

The lemma gives an integer n and a k-uniform hypergraph $\mathrm{H}_{st+1} = (V(\mathrm{H}_{st+1}), E(\mathrm{H}_{st+1}))$ such that

(i) $V(\mathrm{H}_{st+1}) = V_1 \cup V_2 \cup \ldots \cup V_{st+1}$, with $|V_i| = n$ for all i;
(ii) for $\alpha = 1, 2$, for all $1 \le j_1 < j_2 < \ldots < j_{c_\alpha} \le st + 1$, for all $V'_{j_\beta} \subseteq V_{j_\beta}$ with $|V'_{j_\beta}| \ge \epsilon n$ for $\beta = 1, 2, \ldots, c_\alpha$, there exists $e \in E(\mathrm{H}_{st+1})$ such that $e \subseteq \bigcup_{\beta=1}^{\alpha} V'_{j_\beta}$;
(iii) for all $e \in E(\mathrm{H}_{st+1})$, $e \cap V_j \ne \emptyset$ for exactly q or k distinct indices j; and,
(iv) $\mathrm{girth}(\mathrm{H}_{st+1}) > l$.

Observe that $\chi^{(q)}(\mathrm{H}_{st+1}) \le st + 1$ because, by (iii), the classes V_i provide a colouring for which each edge receives at least q colours.

Suppose that $V(\mathrm{H}_{st+1}) = C_1 \cup C_2 \cup \ldots \cup C_{st}$ is a partition. For each $j = 1, 2, \ldots, st + 1$ choose $i(j)$ from $i = 1, 2, \ldots, st$ such that $|V_j \cap C_{i(j)}|$ is maximum among all $|V_j \cap C_i|$. Then $|V_j \cap C_{i(j)}| \ge n/st > \epsilon n$ and, by the pigeonhole principle, $i(1) = i(2)$, say. By (ii), with $\alpha = 2$ and $\vec{k}^{(2)}$ as above, there exists $e \in E(\mathrm{H}_{st+1})$ with

$$e \subseteq \bigcup_{j=1}^{q} V_j \cap C_{i(j)} \subseteq C_{i(1)} \cup C_{i(3)} \cup \ldots \cup C_{i(q)}.$$

Thus, no partition of $V(\mathrm{H}_{st+1})$ into st or fewer parts can provide a colouring of H_{st+1} in which each edge gets at least q colours.

We have shown that $\chi^{(q)}(\mathrm{H}_{st+1}) = st + 1$.

We now prove that H_{st+1} is not s-partition p-homomorphic to A. Let us argue that for an arbitrary partition $V(\mathrm{H}_{st+1}) = X_1 \cup X_2 \cup \ldots \cup X_s$, and $\mathrm{H}^{(i)} = \mathrm{H}[X_i]$ $(i = 1, 2, \ldots, s)$, there is some i such that $\mathrm{H}^{(i)}$ is not p-homomorphic to A. For each $j = 1, 2, \ldots, st + 1$ choose $i(j)$ from $i = 1, 2, \ldots, s$ such that $|V_j \cap X_{i(j)}|$ is maximum among all $|V_j \cap X_i|$. In particular, $|V_j \cap X_{i(j)}| \ge n/s$. By the pigeonhole principle, we may assume, without loss of generality, that

$$X_{i(1)} = X_{i(2)} = \cdots = X_{i(t+1)} = X.$$

Let $B_j = V_j \cap X$, $j = 1, 2, \ldots, t + 1$, and let $\mathrm{H}' = \mathrm{H}_{st+1}[B_1 \cup B_2 \cup \ldots \cup B_{t+1}]$. We shall show that H' is not p-homomorphic to A.

Assume to the contrary that f is a p-homomorphism of H' to A. For $j = 1, 2, \ldots, t + 1$, choose $a_j \in V(\mathrm{A})$ such that $|f^{-1}(a_j) \cap B_j|$ is maximum among all $|f^{-1}(a) \cap B_j|$, $a \in V$. With $U_j = f^{-1}(a_j) \cap B_j$, we see that

$$|U_j| \geq \frac{|B_j|}{|A|} \geq \frac{n/s}{|A|} \geq \epsilon n,\ j = 1, 2, \ldots, t+1.$$

Let $a_1, a_2, \ldots, a_v$ be the distinct elements among $a_1, a_2, \ldots a_{t+1}$.

We first argue that $v \leq \omega$. This is trivial if $v < k$ since $k \leq \omega$. So, suppose that $v \geq k$. We claim that for each k-set $\{i_1, i_2, \ldots, i_k\} \subseteq [v]$, $\{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\} \in E = E(\mathrm{A})$. From (ii), using $\alpha = 1$ and $\vec{k}^{(1)}$ as given, there is $e \in E(\mathrm{H}_{st+1})$ such that $e \subseteq \bigcup_{\beta=1}^{k} U_{i_\beta}$. It is clear that $e \in E(\mathrm{H}')$. Thus,

$$f[e] = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\} \text{ and } f[e] \subseteq e' \in E,$$

so $e' = \{a_{i_1}, a_{i_2}, \ldots, a_{i_k}\} \in E$. Thus, $\{a_1, a_2, \ldots, a_v\}$ induces a clique in A, so $v \leq \omega$.

We note that $v \geq p$. This is because $t + 1 \geq q$, so the subhypergraph of H' induced by the union of the $U_j{}'s$ is nonempty and, so, the image of an edge under the p-homomorphism f must contain at least p distinct $a_i{}'s$. For $i = 1, 2, \ldots, v$, let β_i be the number of j's such that $f[U_j] = \{a_i\}$, and assume that $\beta_1 \geq \beta_2 \geq \ldots \geq \beta_v$. We now consider the size of $\beta = \beta_1 + \beta_2 + \cdots + \beta_{p-1}$. If $\beta \geq q$ then there are distinct indices $j_1, j_2, \ldots, j_q$ such that each $f[U_{j_i}] \subseteq \{a_1, a_2, \ldots, a_{p-1}\}$. By (ii), with $\alpha = 2$ and $\vec{k}^{(2)}$ as given, $\bigcup_{i=1}^{q} U_{j_i}$ contains an edge of H_{st+1} and, hence, of H'. This contradicts the assumption that f is a p-homomorphism. On the other hand, if

$$\beta \leq q - 1 = d(p-1) + r - 1,$$

then, using the fact that $\beta/(p-1) \geq \beta_p \geq \ldots \geq \beta_v$, we have

$$\frac{1}{p-1}\left(d(p-1) + r - 1\right) \geq \beta_p \geq \ldots \geq \beta_v.$$

Since each β_i is an integer and $r \leq p - 1$, this implies that $d \geq \beta_p \geq \ldots \geq \beta_v$. This gives a contradiction as follows:

$$\begin{aligned} t + 1 &= d\omega + r \\ &= \beta_1 + \beta_2 + \ldots + \beta_v \\ &\leq d(p-1) + r - 1 + (v - p + 1)d \\ &\leq d\omega + r - 1 \\ &= t \end{aligned} \tag{1}$$

Again, we see that the assumption that f is a p-homomorphism leads to a contradiction.

Proof of Theorem 2.1 (2)

Recall that $\mathrm{A} = (V, E)$ is a k-uniform hypergraph, and that p, q are integers satisfying $1 < q < p \leq k$. We wish to show that $\psi_{p,q,s}(\mathrm{A}) = s(q-1) + 1$.

To see that $\psi_{p,q,s}(\mathrm{A}) > s(q-1)$, we let H be a k-uniform hypergraph with $\chi^{(q)}(\mathrm{H}) = s(q-1)$ and assume that $C_1, C_2, \ldots, C_{s(q-1)}$ are classes of a q-colouring of H. Define

$$X_j = C_{(j-1)(q-1)+1} \cup C_{(j-1)(q-1)+2} \cup \ldots \cup C_{j(q-1)},\ j = 1, 2, \ldots, s.$$

Since each edge of H must intersect at least q C_j's, each of the subhypergraphs $\mathrm{H}[X_i]$ is empty. Hence, we have p-homomorphisms to A, trivially. This shows that $\psi_{p,q,s}(\mathrm{A}) \geq s(q-1) + 1$.

To see the opposite inequality, we invoke the probabilistic lemma to obtain the required k-uniform hypergraph with girth greater than a specified l. Again, we must give m and h, the value of ϵ and the k-sequences $\vec{k}^{(\alpha)} = (k_1^{(\alpha)}, k_2^{(\alpha)}, \ldots, k_{c_\alpha}^{(\alpha)})$, $\alpha = 1, 2, \ldots, h$.

Let $m = s(q-1)$, $h = 1$, $\epsilon = 1/(s(q-1)|V|)$, and $\vec{k}^{(1)} = (k-q+1, 1, \ldots, 1)$. The lemma gives an integer n and a k-uniform hypergraph $\mathrm{H}_{s(q-1)+1} = (V(\mathrm{H}_{s(q-1)+1}), E(\mathrm{H}_{s(q-1)+1}))$ such that

(i) $V(\mathrm{H}_{s(q-1)+1}) = V_1 \cup V_2 \cup \ldots \cup V_{s(q-1)+1}$, with $|V_i| = n$ for all i;
(ii) for all $1 \leq j_1 < j_2 < \ldots < j_q \leq s(q-1)+1$, for all $V'_{j_\beta} \subseteq V_{j_\beta}$ with $|V'_{j_\beta}| \geq \epsilon n$ for $\beta = 1, 2, \ldots, q$, there exists $e \in E(\mathrm{H}_{s(q-1)+1})$ such that $e \subseteq \bigcup_{\beta=1}^{q} V'_{j_\beta}$;
(iii) for all $e \in E(\mathrm{H}_{s(q-1)+1})$, $e \cap V_j \neq \emptyset$ for exactly q indices j; and,
(iv) $\mathrm{girth}(\mathrm{H}_{s(q-1)+1}) > l$.

Observe that $\chi^{(q)}(\mathrm{H}_{s(q-1)+1}) \leq s(q-1)+1$ because, by (iii), the classes V_i provide a q-colouring. The argument that $\chi^{(q)}(\mathrm{H}_{s(q-1)+1}) > s(q-1)$ is almost exactly the same as that showing $\chi^{(q)}(\mathrm{H}_{st+1}) > st$ in the previous part, and so we omit it.

Let us show that for every s-partition of $V(\mathrm{H}_{s(q-1)+1})$ there is some part that induces a subhypergraph of $\mathrm{H}_{s(q-1)+1}$ with no p-homomorphism to A. Suppose that $V(\mathrm{H}_{s(q-1)+1}) = X_1 \cup X_2 \cup \ldots \cup X_s$ is any partition. Use the pigeonhole principle, just as in the previous part, to show that there are q distinct indices $j_1, j_2, \ldots, j_q$ in $[s(q-1)+1]$ and some $X = X_j$ such that

$$U_i = V_{j_i} \cap X \text{ satisfies } |U_i| \geq \frac{n}{s} > \epsilon n \quad i = 1, 2, \ldots, q.$$

Let $\mathrm{H}' = \mathrm{H}_{s(q-1)+1}[X]$ and let's consider any map f of H' to A. We choose $a_i \in V$ so that $|f^{-1}(a_i) \cap U_i|$ is maximum among all $|f^{-1}(a) \cap U_i|$, $a \in V$, for $i = 1, 2, \ldots, q$. Because $|f^{-1}(a_i) \cap U_i| \geq n/(s|A|)$, (ii) yields $e \in E(\mathrm{H}')$ such that

$$e \subseteq \bigcup_{i=1}^{q} f^{-1}(a_i) \cap U_i, \text{ implying } f[e] \subseteq \{a_1, a_2, \ldots, a_q\}.$$

Since $q < p$, f is not a p-homomorphism. Hence, $\mathrm{H}_{s(q-1)+1}$ is not s-partition p-homomorphic to A.

4 The Proof of the Probabilistic Lemma

We shall construct H by deleting edges from a union $\bigcup_{\alpha=1}^{h} \mathrm{H}^{(\alpha)}$, where each $\mathrm{H}^{(\alpha)}$ corresponds to some k-sequence $\vec{k}^{(\alpha)}$, $\alpha = 1, 2, \ldots, h$. Our probabilistic proof is a modification of the well-known argument due to Erdős [Erd59] and Erdős and Hajnal [EH66].

First, here are the two basic results from probability that we shall need. Following notation in [JŁR00], $\mathbf{X} \in \mathrm{Bi}(r, p)$ means that $\mathbf{X}$ is a random variable with binomial distribution, the sum of r independent Bernoulli random variables. The statement of Chernoff's Inequality is based on Theorem 2.1 in [JŁR00].

Markov's Inequality : for $\mathbf{X} \geq 0$ and $t > 0$,

$$\mathbb{P}(\mathbf{X} \geq t) \leq \frac{\mathbb{E}(\mathbf{X})}{t} .$$

Chernoff's Inequality : for $\mathbf{X} \in \mathrm{Bi}(r, p)$ and $0 \leq \delta \leq 1$,

$$\mathbb{P}(\mathbf{X} \leq \delta r p) \leq \exp\left(-\frac{(1-\delta)^2}{2} r p\right) .$$

Proof of Lemma 2.2

Let $V_1, V_2, \ldots, V_{m+1}$ be disjoint sets each of size n, with n sufficiently large. To construct $\mathrm{H}^{(\alpha)}$, $\alpha = 1, 2, \ldots, h$, we shall select k-element sets, each with probability $p = (\log n)/n^{k-1}$ from the set

$$\bigcup \left(\bigcup_{i=1}^{c_\alpha} \binom{V_{j_i}}{k_i^{(\alpha)}} : \ 1 \leq j_1 < j_2 < \ldots < j_{c_\alpha} \leq m+1 \right) .$$

Thus, there are

$$\binom{m+1}{c_\alpha} \prod_{i=1}^{c_\alpha} \binom{n}{k_i^{(\alpha)}}$$

independent trials in total.

Since our eventual hypergraph H is obtained by deleting edges from the union of the $\mathrm{H}^{(\alpha)}$'s, we see that **(1)** of the lemma is immediate.

Second, we consider property **(2)** of the lemma. For an arbitrary $1 \leq j_1 < j_2 < \ldots < j_{c_\alpha} \leq m+1$, choose a sequence of c_α sets $V'_{j_\beta} \subseteq V_{j_\beta}$, with $|V'_{j_\beta}| \geq \epsilon n$. We call such a sequence *large*. Let $\mathbf{X}(V'_{j_1}, V'_{j_2}, \ldots, V'_{j_{c_\alpha}})$ denote the random variable counting the number of k-sets e satisfying

$$|e \cap V'_{j_\beta}| = k_\beta, \ \beta = 1, 2, \ldots, c_\alpha. \tag{2}$$

Then $\mathbf{X} \in \mathrm{Bi}(r, p)$ where

$$r = \prod_{\beta=1}^{c_\alpha} \binom{|V_{j_\beta}|}{k_\beta} \quad \text{and} \quad p = \frac{\log n}{n^{k-1}}.$$

Thus,

$$\mathbb{E}(\mathbf{X}) = rp \geq \prod_{\beta=1}^{c_\alpha} \left(\frac{\epsilon n}{k_\beta}\right)^{k_\beta} \frac{\log n}{n^{k-1}} \geq \left(\frac{\epsilon}{k}\right)^k n \log n.$$

Apply Chernoff's inequality with $\delta = 1/2$ to obtain

$$\mathbb{P}\left(\mathbf{X} \leq \frac{1}{2}\left(\frac{\epsilon}{k}\right)^k n \log n\right) \leq \exp\left(-\frac{1}{8}\left(\frac{\epsilon}{k}\right)^k n \log n\right).$$

The number of sequences $(V_{j_1}^{'}, V_{j_2}^{'} \dots, V_{j_{c_\alpha}}^{'})$ with $1 \leq j_1 < j_2 < \dots < j_{c_\alpha}$ is less than $\binom{m+1}{c_\alpha} 2^{c_\alpha n}$ and thus the probability that there is a large sequence which contains fewer than $\frac{1}{2}\left(\frac{\epsilon}{k}\right)^k n \log n$ satisfying (2) is at most

$$\binom{m+1}{c_\alpha} 2^{c_\alpha n} \exp\left(-\frac{1}{8}\left(\frac{\epsilon}{k}\right)^k n \log n\right) \tag{3}$$

which tends to 0 as $n \to \infty$.

We now want to estimate the number of circuits of length at most r in the hypergraph $\mathrm{H} = \bigcup_{\alpha=1}^{h} \mathrm{H}^{(\alpha)}$. First, recall that a circuit of length $j \geq 2$ corresponds to a family of distinct k-sets $e_1, e_2, \dots, e_j$ such that $|\bigcup_{i=1}^{j} e_i| \leq j(k-1)$ and, conversely, a family satisfying these conditions contains a circuit of length at most j. The number of such families is less than

$$\binom{(m+1)n}{j(k-1)} \binom{j(k-1)}{k}^j \leq d_j(m,k) n^{j(k-1)} ,$$

where $d_j(m,k) = d_j$ depends only on j, k, m.

Let $\mathbf{Y}_j$ be the random variable counting the number of j-circuits $j = 1, 2, \dots, l$. Then

$$\mathbb{E}(\mathbf{Y}_j) \leq d_j n^{j(k-1)} p^j = d_j \log^j n .$$

Let $\mathbf{Y} = \sum_{j=2}^{l} \mathbf{Y}_j$. Then $\mathbb{E}(\mathbf{Y}) \leq \sum_{j=2}^{l} d_j \log^j n < d_0 \log^l n$. Apply Markov's inequality to $\mathbb{E}(\mathbf{Y})$:

$$\mathbb{P}(\mathbf{Y} > 2d_0 \log^l n) \leq \frac{\mathbb{E}(\mathbf{Y})}{2d_0 \log^l n} = \frac{1}{2} . \tag{4}$$

Summarizing (3) and (4), we infer that for $n \geq n(m,k,l,\epsilon)$, there is a k-uniform hypergraph H_n on the vertex set $\bigcup_{i=1}^{m+1} V_i$ such that

(i) for all $\alpha = 1, 2, \dots, h$, $1 \leq j_1 < j_2 < \dots < j_{c_\alpha} \leq m+1$ and large sequences $(V_{j_1}^{'}, V_{j_2}^{'}, \dots, V_{j_{c_\alpha}}^{'})$ there exist at least $\frac{1}{2}\left(\frac{\epsilon}{k}\right)^k n \log n$ edges $e \in \mathrm{H}^{(\alpha)}$ such that $|e \cap V_{j_\beta}^{'}| = k_\beta$, for $\beta = 1, 2, \dots, c_\alpha$; and,

(ii) the number of circuits of length at most l is at most $2d_0\log^l n$.

Delete one edge from each such circuit, resulting in a k-uniform hypergraph H of girth greater than l in which each large family contains at least one edge of H. Thus, we have item **(3)** of the lemma.

This completes the proof of Lemma 2.2.

5 A Packing Construction

An alternative approach to Theorem 1.4 uses Theorem 1.1 and the well-known result of Erdős to establish the theorem for $s = 1$ [see Lemma 5.1]. Then a packing construction for hypergraphs [Lemma 5.2] and a probabilistic construction [Lemma 5.3] can be used to both generalize from the graph case to hypergraphs and extend the result to s-partitions. This argument may be of independent interest, so we present it in this section.

We revert to the usual meanings of homomorphism, colouring, and chromatic number for graphs and hypergraphs, as used in Section 1. That is, a homomorphism of graphs is a 2-homomorphism, of k-uniform hypergraphs, a k-homomorphism, a colouring c of a hypergraph H has $|c[e]| \geq 2$, for all edges e of H, and chromatic number for hypergraphs is the [weak] chromatic number.

First, here is Theorem 1.4 in the case $s = 1$, proved using probabilistic results from [Erd59] and [NZ04].

Lemma 5.1. *Let* A *be a k-uniform hypergraph and let $l \geq 2$. Then there is a k-uniform hypergraph* C *with* $\mathrm{girth}(\mathrm{C}) > l$ *and* $\chi(\mathrm{C}) = \psi(\mathrm{A})$ *such that $C \nrightarrow A$.*

Proof. Let A be a graph with $\omega(\mathrm{A}) = \omega$ and let $l \geq 2$. As noted in the introduction, it is straightforward to argue that

$$\psi(\mathrm{A}) = \begin{cases} \omega(\mathrm{A}) + 1 & \text{if } k = 2, \\ 2 & \text{if } k \geq 3, \end{cases}$$

if there is no restriction on girth.

Let C be a graph with $\mathrm{girth}(\mathrm{C}) > l$, obtained from Theorem 1.1 in the case that $\mathrm{H} = \mathrm{K}_{\omega+1}$ and $n = \max\{|V(\mathrm{A})|, \omega + 1\}$. Since $\mathrm{K}_{\omega+1} \nrightarrow \mathrm{A}$, we have that $\mathrm{C} \nrightarrow \mathrm{A}$. On the other hand, $\mathrm{C} \to \mathrm{K}_{\omega+1}$, and thus $\chi(\mathrm{C}) \leq \omega + 1$. Of course, $\mathrm{C} \nrightarrow \mathrm{K}_\omega$ because $\mathrm{K}_{\omega+1} \nrightarrow \mathrm{K}_\omega$. Thus, $\chi(\mathrm{G}) = \omega + 1$. This finishes the argument in the case of graphs.

Let $k \geq 3$ and let A be a k-uniform hypergraph. Given l, we construct a k-uniform hypergraph C such that $\chi(\mathrm{C}) = 2$, $\mathrm{girth}(\mathrm{C}) > l$, and $\mathrm{C} \nrightarrow \mathrm{A}$.

For a k-uniform hypergraph H, we denote by $Gr(\mathrm{H})$ the graph on vertex set $V(\mathrm{H})$, with edge set all pairs $\{u, v\}$ such that u and v belong to an edge of H. [This is a special case of the construction below and is called the *2-section* of H in [Ber89].] Note that if f is a homomorphism of H to a k-uniform hypergraph L then f is a graph homomorphism of $Gr(\mathrm{H})$ to $Gr(\mathrm{L})$.

Let $\chi(Gr(\mathrm{A})) = c$. By the well-known result in [Erd59], there exists a graph G such that $\chi(\mathrm{G}) \geq c+1$ and girth(G) $> l$. Let W be a set of size $(k-2)|E(\mathrm{G})|$ which is disjoint from $V(\mathrm{G})$. We associate with every edge e of G a subset $\delta(e)$ of W with $|\delta(e)| = k-2$ and $\delta(e) \cap \delta(f) = \emptyset$ for any two different edges e and f of G.

Let the k-uniform hypergraph C have $V(\mathrm{C}) = V(\mathrm{G}) \cup W$ and the set of edges

$$\{e \cup \delta(e) : e \text{ is an edge of G}\}.$$

Since $k \geq 3$ the partition $V(\mathrm{G}) \cup W$ is a 2-colouring of C. And, $\mathrm{C} \nrightarrow \mathrm{A}$ because $\chi(Gr(\mathrm{C})) \geq c+1$, while $\chi(Gr(\mathrm{A})) = c$. It is easy to see that girth(C) $> l$. □

In the rest of this section, we outline the alternative approach to Theorem 1.4 for arbitrary s. This requires a few new terms.

Let A be a k-uniform hypergraph and let H be a simple $|V(\mathrm{A})|$-uniform hypergraph. We let $\mathcal{H}(\mathrm{A})$ denote the set of all k-uniform hypergraphs obtained by inserting a copy of A in each edge of H. More formally, $\widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{A})$ if and only if $V(\widehat{\mathrm{H}}) = V(\mathrm{H})$, for all $e \in E(\mathrm{H})$, $\widehat{\mathrm{H}}[e]$ is isomorphic to A, and for all $\widehat{e} \in E(\widehat{\mathrm{H}})$ there is some $e \in E(\mathrm{H})$ such that $\widehat{e} \subseteq e$. Regarding H as a labelled hypergraph, with the labelling inherited by each member of $\mathcal{H}(\mathrm{A})$, we see that

$$|\mathcal{H}(\mathrm{A})| = \left(\frac{|V(\mathrm{A})|!}{|\mathrm{Aut}(\mathrm{A})|}\right)^{|E(\mathrm{H})|}.$$

Refer to any member of $\mathcal{H}(\mathrm{A})$ as an A-*packing* of H.

Let A be a hypergraph and s be a nonnegative integer. Let

$$\pi(\mathrm{A}, s+1) = \min\{\chi(\widehat{\mathrm{H}}) \ : \ \chi(\mathrm{H}) = s+1, \ \widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{A})\}.$$

with the minimum taken over all $(s+1)$-chromatic $|V(\mathrm{A})|$-uniform hypergraphs H and all $\widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{A})$.

Lemma 5.2. *Let* A *be a hypergraph and let* s *be a nonnegative integer. Then*

$$\pi(\mathrm{A}, s+1) \geq (\chi(\mathrm{A}) - 1)s + 1.$$

Proof. Let H be any simple $|V(A)|$-uniform hypergraph of chromatic number $s+1$ and let ϕ be any mapping of $V(\mathrm{H})$ to $X \times Y$ where $|X| = \chi(\mathrm{A}) - 1$ and $|Y| = s$. For each $v \in V(\mathrm{H})$ let $\hat{\phi}(v) = y$ where $\phi(v) = (x, y)$. Since $|Y| < \chi(\mathrm{H})$, there is some $e \in E(\mathrm{H})$ such that $\hat{\phi}|e$ is constant. Now let $\widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{A})$. Since $|X| < \chi(\mathrm{A})$ and $\widehat{\mathrm{H}}[e]$ is isomorphic to A, there is some $e' \in E(\mathrm{A})$ such that $e' \subseteq e$ and $\phi|e'$ is constant. Thus, ϕ cannot be a colouring of $\widehat{\mathrm{H}}$. Hence, $\pi(\mathrm{A}, s+1) \geq (\chi(\mathrm{A}) - 1)s + 1$. □

Lemma 5.3. *let* A *be a* k*-uniform hypergraph and let* s *and* l *be positive integers. Then there is a* $|V(\mathrm{A})|$*-uniform hypergraph* H *such that* $\chi(\mathrm{H}) = s+1$, girth(H) $> l$, *and an* A*-packing* $\widehat{\mathrm{H}}$ *of* H *such that*

$$\chi(\widehat{\mathrm{H}}) = (\chi(\mathrm{A}) - 1)s + 1.$$

Consequently,

$$\chi(\widehat{\mathrm{H}}) = \pi(\mathrm{A}, s+1) = (\chi(\mathrm{A}) - 1)s + 1.$$

Proof. We are given k, s and l and the k-uniform hypergraph A. Let $|V(\mathrm{A})| = a$, $\chi = \chi(\mathrm{A})$, and $V(\mathrm{A}) = W_1 \cup W_2 \cup \ldots \cup W_\chi$ be a partition induced by a χ-colouring of A, with $|W_i| = a_i$ for $i = 1, 2, \ldots, \chi$.

We apply Lemma 2.2 with $k = a$, $h = 1$, $c_1 = \chi$, $\epsilon = 1/s$, the a-sequence $\vec{a} = (a_1, a_2, \ldots, a_\chi)$, and $m = (\chi - 1)s$. This yields an a-uniform H satisfying **(1)**, **(2)** and **(3)** of the lemma. Define a k-uniform hypergraph $\widehat{\mathrm{H}}$ on $V(\mathrm{H})$ as follows: for each $e \in E(\mathrm{H})$ insert a copy of A in e by identifying W_β with $e \cap V_{j_\beta}$, where $|e \cap V_{j_\beta}| = a_\beta$ for $\beta = 1, 2, \ldots, \chi$, as guaranteed by **(1)**. It is clear that $\widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{A})$.

We know from Lemma 5.2 that $\chi(\widehat{\mathrm{H}}) \geq m + 1$. The upper bound is immediate from the construction of $\widehat{\mathrm{H}}$, so we see that $\chi(\widehat{\mathrm{H}}) = m + 1$. It remains to prove that $\chi(\mathrm{H}) = s + 1$.

To see that $\chi(\mathrm{H}) \leq s + 1$, recall that $V(\mathrm{H}) = \bigcup_{i=1}^{m+1} V_i$. Create a new partition of $V(\mathrm{H})$ with parts U_j, $j = 1, 2, \ldots, s+1$, by letting each U_j be the union of $\chi - 1$ distinct V_i's, for $j = 1, 2, \ldots, s$, and $U_{s+1} = V_{m+1}$. By **(1)** of Lemma 2.2, each edge of H intersects exactly χ V_i's, so the partition by U_j, $j = 1, 2, \ldots, s+1$ provides an $(s+1)$-colouring of H.

To prove that $\chi(\mathrm{H}) > s$, suppose that $V(\mathrm{H}) = X_1 \cup X_2 \cup \ldots \cup X_s$ is a partition. For each $i = 1, 2, \ldots, m+1 = (\chi - 1)s + 1$ there is some $j(i) \in [s]$ such that $|V_i \cap X_{j(i)}| > n/s = \epsilon n$. There is some index j_0 such that $j_0 = j(i)$ for at least χ distinct i's. By **(2)**, X_{j_0} contains an edge of H, so the partition by X_j, $j = 1, 2, \ldots, s$ cannot give an $(s-1)$-colouring of H. □

We can now state and prove a somewhat sharpened version of Theorem 1.4.

Theorem 5.4. *Let* A *be a k-uniform hypergraph and let s and l be positive integers. Then there exist a k-uniform hypergraph* C *and a $|V(\mathrm{C})|$-uniform hypergraph* H*, both of girth greater than l, such that $\chi(\mathrm{C}) = \psi(\mathrm{A})$, $\mathrm{C} \nrightarrow \mathrm{A}$, $\chi(\mathrm{H}) = s + 1$ and such that some $\widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{C})$ satisfies:*

(1) $\widehat{\mathrm{H}}$ *is not s-partition homomorphic to* A*;*
(2) $\chi(\widehat{\mathrm{H}}) = \psi(\mathrm{A}, s) = s(\psi(A) - 1) + 1$*; and*
(3) $\mathrm{girth}(\widehat{\mathrm{H}}) > l$.

Just as in Theorem 1.4, this gives the value of $\psi(\mathrm{A}, s)$, once we recall that $\psi(\mathrm{A}) = \omega(\mathrm{A}) + 1$ for graphs and $\psi(\mathrm{A}) = 2$ for nontrivial hypergraphs, and shows that there is a large girth hypergraph realizing $\psi(\mathrm{A}, s)$.

Proof. Let A, s and l be as in the preceding statement. Apply Lemma 5.1 to A and l to obtain a k-uniform C such that $\mathrm{girth}(\mathrm{C}) > l$, $\chi(\mathrm{C}) = \psi(\mathrm{A})$, and $\mathrm{C} \nrightarrow \mathrm{A}$.

Apply Lemma 5.3 to C, $s+1$ and l to obtain a $|V(C)|$-uniform hypergraph H such that girth(H) $> l$, and $\chi(\mathrm{H}) = s+1$, and, for some $\widehat{\mathrm{H}} \in \mathcal{H}(\mathrm{C})$

$$\chi(\widehat{\mathrm{H}}) = \pi(\mathrm{C}, s+1) = (\chi(\mathrm{C})-1)s+1. \tag{5}$$

For any partition of the vertices of $V(\widehat{\mathrm{H}}) = V(\mathrm{H})$ into s parts, some part, say W, contains an edge of H. Then $\widehat{\mathrm{H}}[W]$ contains a copy of A and, thus, has no homomorphism to A. Therefore, $\widehat{\mathrm{H}}$ is not s-partition homomorphic to A, proving **(1)**.

The proof of **(2)** will be complete once we show that $\psi(\mathrm{A}, s) = \pi(\mathrm{C}, s+1)$. Let us see that

$$\psi(\mathrm{A}, s) \leq \pi(\mathrm{C}, s+1). \tag{6}$$

Given any simple $(s+1)$-chromatic, $|V(\mathrm{C})|$-uniform hypergraph F and any $\widehat{\mathrm{F}} \in \mathcal{F}(\mathrm{C})$, a partition of $\widehat{\mathrm{F}}$ into s parts must result in a part which contains an edge of F and, hence, a copy of C. Then the subhypergraph of $\widehat{\mathrm{F}}$ induced on that part has no homomorphism to A, so $\widehat{\mathrm{F}}$ is not s-partition homomorphic to A and $\psi(\mathrm{A}, s) \leq \chi(\widehat{\mathrm{F}})$. Since F and $\widehat{\mathrm{F}}$were chosen arbitrarily, $\psi(\mathrm{A}, s) \leq \pi(\mathrm{C}, s+1)$.

In order to prove the opposite inequality, we need that

$$s(\psi(\mathrm{A})-1)+1 \leq \psi(\mathrm{A}, s). \tag{7}$$

To prove it, let F be any k-uniform hypergraph with $\chi(\mathrm{F}) \leq s(\psi(\mathrm{A})-1)$. Let V_i $(i = 1, 2, \ldots, s(\psi(\mathrm{A})-1))$ be the classes of a colouring of F. Now let W_j $(j = 1, 2, \ldots, s)$ be pairwise disjoint, with each W_j the union of $\psi(\mathrm{A})-1$ distinct V_i's. Since each of the s induced subhypergraphs $\mathrm{F}[W_j]$ has a $(\psi(\mathrm{A})-1)$-colouring, defined by the V_i's contained in W_j, for $j = 1, 2, \ldots, s$, $\mathrm{F}[W_j] \to \mathrm{A}$. Thus, F is s-partition homomorphic to A. This proves that $s(\psi(\mathrm{A})-1)+1 \leq \psi(\mathrm{A}, s)$.

We now conclude that $\psi(\mathrm{A}, s) = \pi(\mathrm{C}, s+1)$ from (8) below, which follows from (7), (6), (5), and the fact that $\chi(\mathrm{C}) = \psi(\mathrm{A})$:

$$\begin{aligned} s(\psi(\mathrm{A})-1)+1 \leq \psi(\mathrm{A}, s) \leq \pi(\mathrm{C}, s+1) \leq \\ \leq s(\chi(\mathrm{C})-1)+1 \leq s(\psi(\mathrm{A})-1)+1. \end{aligned} \tag{8}$$

Finally, **(3)**, regarding the girth of a packing, is an immediate consequence of the following observation: girth$(\widehat{\mathrm{H}}) \geq \min\{\mathrm{girth(H)}, \mathrm{girth(C)}\}$. □

6 Explicit Packings in Specific Cases

Given s and a k-uniform hypergraph A, Theorem 5.4 produces an $(s+1)$-chromatic, large girth $|V(\mathrm{A})|$-uniform hypergraph H such that some A-packing $\widehat{\mathrm{H}}$ of H has chromatic number equal to the minimum possible, $\pi(\mathrm{A}, s+1) = (\chi(\mathrm{A})-1)s+1$.

The methods used are probabilistic, so there is no explicit description of $\widehat{\mathrm{H}}$. We present an explicit construction of large girth hypergraphs for $\mathrm{A} = \mathrm{K}_k$. Here is a restatement of the objective: for integers k, s and l, explicitly construct a k-uniform hypergraph H of girth greater than l, weak chromatic number $s+1$, and strong chromatic number $(k-1)s+1$.

The construction is based on a general method created by Nešetřil and Rödl [NR79]. The following is not explicitly stated in [NR79] but is a consequence of the proof of the main theorem of that paper.

Lemma 6.1 (The Girth Machine). *Let* L *be a k-uniform hypergraph with $V(\mathrm{L}) = \{v_1, v_2, \ldots, v_t\}$, weak chromatic number χ_w, and strong chromatic number χ_s. Then there exists a k-uniform hypergraph $\widetilde{\mathrm{L}}$ with $V(\widetilde{\mathrm{L}}) = V_1 \cup V_2 \cup \ldots \cup V_t$, weak chromatic number χ_w, strong chromatic number χ_s, and* $\mathrm{girth}(\widetilde{\mathrm{L}}) > l$*. Moreover, the mapping $V_i \to v_i$ is a homomorphism of $\widetilde{\mathrm{L}}$ to* L.

Apply the Girth Machine to the k-uniform hypergraph $L = \mathrm{K}^{(k)}_{(k-1)s+1}$, that is, the hypergraph on vertex set $[(k-1)s+1]$ with edge set all k-element subsets of $[(k-1)s+1]$. The weak chromatic number of $\mathrm{K}^{(k)}_{(k-1)s+1}$ is $s+1$, by a pigeonhole argument, and the strong chromatic number is the size of the vertex set, $(k-1)s+1$. Therefore, $\widetilde{\mathrm{L}}$ has the desired properties. Unfortunately, the cardinality of $V(\widetilde{\mathrm{L}})$ is a tower function of considerable height.

We end with a concrete, small construction that handles the case $k = l = s = 3$.

The set of vertices of H is the set $S := \{1, 2, 3, 4, 5\}$ together with the 10 2-element subsets of S, so $|V(\mathrm{H})| = 15$. The set of hyperedges of H is the set

$$\{\{x, y, \{x, y\}\} : x, y \in S\} \cup \{\{\{x, y\}, \{y, z\}, \{z, x\}\} : x, y, z \in S\}.$$

To see that $\chi(\mathrm{H}) \geq 3$, assume for a contradiction that H has a colouring γ with colours a and b. Then three of the elements in S receive the same colour. We may assume, without loss of generality, that the elements 1, 2 and 3 all receive colour a. Then the three 2-element subsets $\{1, 2\}, \{2, 3\}, \{3, 1\}$ must be coloured with b, which is a contradiction.

Note that the strong chromatic number of H is equal to the total chromatic number of the complete graph K_5, which is 5. (The set {1, {2,5}, {3,4}} is a colour class and one obtains the others by rotation.)

References

[Ber89] C. Berge, *Hypergraphs*, North-Holland Mathematical Library, **45**, Elsevier Science Publishers B.V. (1989).

[Erd59] P. Erdős, Graph theory and probability, *Canad. J. Math.* **11** (1959), 34–38.

[EH66] P. Erdős and A. Hajnal, On chromatic number of graphs and setsystems, *Acta. Math. Acad. Sci. Hung.* **17** (1966), 61–99.

[HN04] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Clarendon Press, Oxford, 2004.

[JŁR00] S. Janson, T. Łuczak and A. Ruciński, *Random Graphs*, Wiley, New York, 2000.

[NR79] J. Nešetřil and V. Rödl, A short proof of the existence of highly chromatic hypergraphs without short cycles, *J Combinatorial Theory B* **27** (1979), 225–227.

[NZ04] J. Nešetřil and X. Zhu, On sparse graphs with given colorings and homomorphisms, *J Combinatorial Theory B* **90** (2004), 161–172.

Acyclic 4-Choosability of Planar Graphs Without Cycles of Specific Lengths

Mickaël Montassier, André Raspaud, and Weifan Wang*

LaBRI UMR CNRS 5800, Université Bordeaux I, 33405 Talence Cedex, France
{montassi,raspaud,weifan}@labri.fr

Summary. A proper vertex coloring of a graph $G = (V, E)$ is acyclic if G contains no bicolored cycle. A graph G is acyclically L-list colorable if for a given list assignment $L = \{L(v) : v \in V\}$, there exists a proper acyclic coloring c of G such that $c(v) \in L(v)$ for all $v \in V$. If G is acyclically L-list colorable for any list assignment with $|L(v)| \geq k$ for all $v \in V$, then G is acyclically k-choosable.

Let G be a planar graph without 4-cycles and 5-cycles. In this paper, we prove that G is acyclically 4-choosable if G furthermore satisfies one of the following conditions: (1) without 6-cycles; (2) without 7-cycles; (3) without intersecting triangles.

AMS Subject Classification. 05C15.
Keywords. Graph coloring, choosability, acyclic.

1 Introduction

Let G be a graph with the vertex set $V(G)$ and the edge set $E(G)$. A proper vertex coloring of G is an assignment c of integers (or labels) to the vertices of G such that $c(u) \neq c(v)$ if the vertices u and v are adjacent in G. A *k-coloring* is a proper vertex coloring using k colors. A proper vertex coloring of a graph is *acyclic* if there is no bicolored cycle in G. The *acyclic chromatic number*, denoted by $\chi_a(G)$, of a graph G is the smallest integer k such that G has an acyclic k-coloring.

The acyclic coloring of graphs were introduced by Grünbaum in [Grü73] and studied by Mitchem [Mit74], Albertson, Berman [AB77], and Kostochka [Kos76]. In 1979, Borodin proved Grünbaum's conjecture:

Theorem 1.1 ([Bor79]). *Every planar graph is acyclically 5-colorable.*

* On leave of absence from the Departement of Mathematics, Zhejiang Normal University, Jinhua 321004, P. R. CHINA. Supported by the french CNRS.

This bound is best possible. In 1973, Grünbaum gave an example of a 4-regular planar graph [Grü73] which is not acyclically colorable with four colors. Moreover, there exist bipartite 2-degenerate planar graphs which are not acyclically 4-colorable [KM76].

Borodin, Kostochka and Woodall improved this bound for planar graphs with a given girth. We recall that the girth of a graph is the length of its shortest cycle.

Theorem 1.2 ([BKW99]).

1. *Every planar graph with girth at least* 7 *is acyclically* 3*-colorable.*
2. *Every planar graph with girth at least* 5 *is acyclically* 4*-colorable.*

A graph G is acyclically L-list colorable if for a given list assignment $L = \{L(v) : v \in V(G)\}$ there is an acyclic coloring c of the vertices such that $c(v) \in L(v)$. We say that c is an L-coloring of G. If G is acyclically L-list colorable for any list assignment with $|L(v)| \geq k$ for all $v \in V(G)$, then G is acyclically k-choosable. The *acyclic list chromatic number* of G, $\chi_a^l(G)$, is the smallest integer k such that G is acyclically k-choosable.

In [BFKRS02], the following theorem is proved and the next conjecture is given:

Theorem 1.3 ([BFKRS02]). *Every planar graph is acyclically* 7*-choosable.*

This means that for any given list assignment L, with $|L(v)| \geq 7$ for all $v \in V(G)$, there is an acyclic coloring c of G, such that it is possible to choose for each vertex v a color in $L(v)$.

Conjecture 1.4 ([BFKRS02]). *Every planar graph is acyclically* 5*-choosable.*

In [MOR05], the authors studied the acyclic choosability of graphs with bounded maximum average degree. The maximum average degree, $Mad(G)$, of the graph G is defined as

$$Mad(G) = \max\{2|E(H)|/|V(H)|, H \subseteq G\}.$$

Theorem 1.5 ([MOR05]).

1. *Every graph* G *with* $Mad(G) < \frac{8}{3}$ *is acyclically* 3*-choosable.*
2. *Every graph* G *with* $Mad(G) < \frac{19}{6}$ *is acyclically* 4*-choosable.*
3. *Every graph* G *with* $Mad(G) < \frac{24}{7}$ *is acyclically* 5*-choosable.*

If G is a planar graph with girth g, then $Mad(G) < \frac{2g}{g-2}$. The following are the immediate consequences of Theorem 1.5.

Corollary 1.6.

1. *Every planar graph with girth at least* 8 *is acyclically* 3*-choosable.*
2. *Every planar graph with girth at least* 6 *is acyclically* 4*-choosable.*
3. *Every planar graph with girth at least* 5 *is acyclically* 5*-choosable.*

In this paper, we focus on the acyclic choosability of planar graph without cycles of specific length. In precise, we prove the following three results. We say that two triangles are *intersecting* if they have at least one vertex in common.

Theorem 1.7. *Every planar graph without cycles of length* 4 *to* 6 *is acyclically* 4*-choosable.*

Theorem 1.8. *Every planar graph without cycles of length* 4, 5, 7 *is acyclically* 4*-choosable.*

Theorem 1.9. *Every planar graph without* 4*-cycles,* 5*-cycles and intersecting triangles is acyclically* 4*-choosable.*

Let G be a plane graph. We use $F(G)$ to denote the set of faces of G. Let $d(v)$ denote the degree of a vertex v in G and $r(f)$ the degree of a face f in G. A *k-vertex* (or *k-face*) is a vertex (or face) of degree k. A k-face having the boundary vertices $x_1, x_2, \cdots, x_k$ in the cyclic order is denoted by $[x_1 x_2 \cdots x_k]$. For a vertex $v \in V(G)$, let $n_i(v)$ denote the number of i-vertices adjacent to v for $i \geq 1$, and $m_3(v)$ the number of 3-faces incident to v. For a face $f \in F(G)$, let $n_i(f)$ denote the number of i-vertices incident to f for $i \geq 2$, and $m_3(f)$ the number of 3-faces adjacent to f. If a 3-face $[uvw]$ satisfies $d(u) = d(v) = 3$ and $d(w) \geq 5$, we call both u and v *light* 3-vertices of G. A 3-face $[xyz]$ of G is called *light* if $d(x) = 3$, $d(y) = d(z) = 4$, and $m_3(y) = m_3(z) = 2$. Suppose that v is a 4-vertex incident to two non-adjacent 3-faces f_1 and f_2. We say that f_1 and f_2 are *opposite* 3-faces with respect to the vertex v. If a vertex v is adjacent to a 3-vertex u such that the edge uv is not incident to any 3-face, then we call u a *pendant* 3-vertex of v.

2 Structural Properties

The proofs of Theorems 1.7 to 1.9 are proved by contradiction. In each case, we suppose that H is a minimal counterexample (i.e., with the smallest vertex number) embedded in the plane to the theorem. Thus G is connected. We first investigate the structural properties of H, then use Euler's formula and the discharging technique to derive a contradiction.

Lemma 2.1. *The minimal counterexample H to Theorems 1.7 to 1.9 satisfies the following.*

(C1) *There are no* 1*-vertices.*
(C2) *A* 2*-vertex is not incident to a* 3*-face.*
(C3) *A* 2*-vertex is not adjacent to a vertex of degree at most* 3.
(C4) *A* 3*-vertex is adjacent to at most one* 3*-vertex.*
(C5) *A* 4*-vertex is adjacent to at most one* 2*-vertex.*
(C6) *There is no* 3*-face incident to two* 3*-vertices and a* 4*-vertex.*
(C7) *H does not contain G_0 as a subgraph.*

(C8) *H does not contain* G_1 *as a subgraph.*
(C9) *H does not contain* G_2 *as a subgraph.*
(C10) *H does not contain* G_3 *as a subgraph.*
(C11) *H does not contain* G_4 *as a subgraph.*
(C12) *H does not contain* G_5 *as a subgraph.*
(C13) *A* 5*-vertex is adjacent to at most three* 2*-vertices.*
(C14) *There is no* 5*-vertex incident to a* 3*-face, adjacent to three* 2*-vertices.*

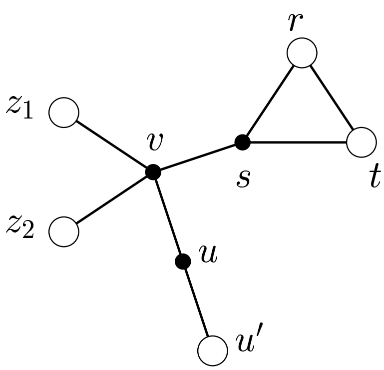


Fig. 1. The reducible subgraph G_0

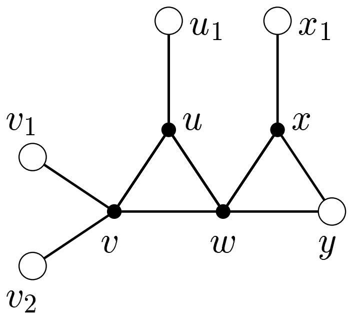


Fig. 2. The reducible subgraph G_1

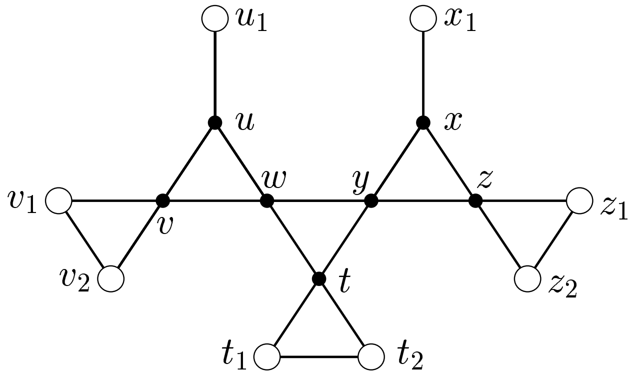


Fig. 3. The reducible subgraph G_2

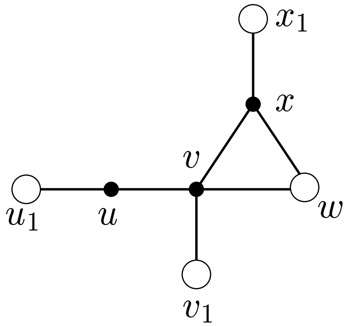


Fig. 4. The reducible subgraph G_3

Proof.

1. Trivial.
2. Suppose that H contains a 2-vertex u incident to a 3-face $[uvw]$. By minimality of H, the graph $H' = H \setminus \{u\}$ is acyclically 4-choosable. Consequently, there exists an acyclic L-coloring c of H'. We can extend this coloring to H by coloring u with $c(u) \in L(u) \setminus \{c(v), c(w)\}$. The coloring obtained is an acyclic L-coloring of H.
3. Suppose that H contains a 2-vertex v adjacent to the vertices u and w such that $d(w) \leq 3$. Let w_1 (and w_2 if $d(w) = 3$) be the other adjacent neighbour of w. By minimality of H, the graph $H' = H \setminus \{v\}$ is acyclically 4-choosable. Let c be an acyclic L-coloring of H'. We can extend this coloring to H:
   3.1. If $c(u) \neq c(w)$, we color v with a color different from these of u and w.
   3.2. If $c(u) = c(w)$, we color v with a color different from these of w, w_1 (and w_2 if $d(w) = 3$).

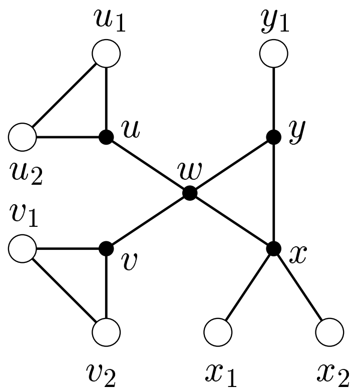


Fig. 5. The reducible subgraph G_4

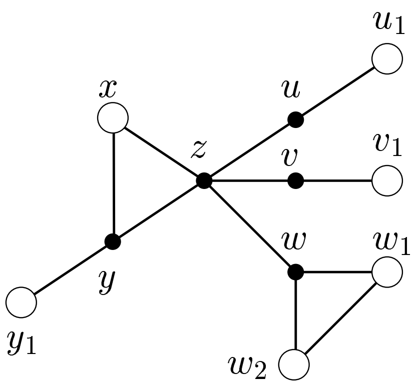


Fig. 6. The reducible subgraph G_5

The obtained coloring is an acyclic L-coloring of H.

4. Suppose that H contains a 3-vertex w adjacent to two 3-vertices v_1, v_2 (each adjacent to u_1, u'_1 and u_2, u'_2) and to another vertex z. Let c be an acyclic L-coloring of $H \setminus \{w\}$. We consider the following cases:
   4.1. $c(v_1)$, $c(v_2)$ and $c(z)$ are pairwise distinct. We color w with a color different from these of v_1, v_2, z.
   4.2. $c(v_1) = c(v_2) \neq c(z)$. W.l.o.g., suppose that $c(v_1) = c(v_2) = 2$ and $c(z) = 1$. Observe that $L(w)$ contains 1 and 2; otherwise, we color w with a color different from 1 or 2 and different from $c(u_1), c(u'_1)$. Assume that $L(w) = \{1, 2, 3, 4\}$. If we cannot color w, this implies that $\{c(u_1), c(u'_1)\} = \{c(u_2), c(u'_2)\} = \{3, 4\}$. As well, observe that $L(v_1) = L(v_2) = \{1, 2, 3, 4\}$; otherwise, we modify the color of v_1

(or v_2) with a color different from 1,2,3,4 to get case 4.1. Hence, we recolor v_1 and v_2 with 1 and color w with 2.

4.3. $c(v_1) = c(z) \neq c(v_2)$. W.l.o.g., suppose that $c(v_1) = c(z) = 1$ and $c(v_2) = 2$. With the same argument as above, we can assume that $L(w) = \{1, 2, 3, 4\}$ and $L(v_1) = \{1, 2, 3, 4\}$. We recolor v_1 with 2 to get case 4.2.

4.4. $c(v_1) = c(v_2) = c(z)$. Observe that $c(u_1) = c(u'_1)$; otherwise, we modify the color of v_1 to get a previous case. We have $c(u_2) = c(u'_2)$ for the same reason and we can choose $c(w) \in L(w) \setminus \{c(u_1), c(u_2), c(z)\}$.

5. Suppose that H contains a 4-vertex v adjacent to two 2-vertices u_1 and u_2 and to two other vertices z_1 and z_2. Let u'_1 and u'_2 be the outer neighbours of u_1 and u_2, respectively. Let c be an acyclic L-coloring of $H \setminus \{u_1\}$. If $c(v) \neq c(u'_1)$, we color u_1 with a color different from $c(v), c(u'_1)$, and the coloring obtained is acyclic. Now, suppose that $c(v) = c(u_1)$. If we cannot color u_1, then w.l.o.g. $L(u_1) = \{1, 2, 3, 4\}$, $c(u'_1) = c(v) = c(u'_2) = 1$, $c(u_2) = 2$, $c(z_1) = 3$, and $c(z_2) = 4$. So, after erasing the color of u_2, we recolor v with a color different from 1, 3, 4, and we color u_i with a color different from $c(v), c(u'_i)$ for $i = 1, 2$. The coloring obtained is acyclic.

6. Suppose that H contains a 3-face $[uvw]$ with $d(u) = d(v) = 3$ and $d(w) = 4$. Let u_1, v_1 and w_1, w_2 be the outer neighbours of u, v and w, respectively. Let c be an acyclic L-coloring of $H \setminus \{u, v\}$. We consider the following cases:

6.1. The colors of u_1, v_1, w are pairwise distinct. We color u with $c(u) \in L(u) \setminus \{c(u_1), c(w)\}$ and v with $c(v) \in L(v) \setminus \{c(u), c(v_1), c(w)\}$.

6.2. $c(u_1) = c(w)$ and $c(v_1) \neq c(w)$. We color u with $c(u) \in L(u) \setminus \{c(w), c(w_1), c(w_2)\}$ and v with $c(v) \in L(v) \setminus \{c(v_1), c(u), c(w)\}$.

6.3. $c(u_1) = c(v_1)$ and $c(w) \neq c(u_1)$. We color u with $c(u) \in L(u) \setminus \{c(u_1), c(w)\}$ and v with $c(v) \in L(v) \setminus \{c(v_1), c(u), c(w)\}$.

6.4. $c(u_1) = c(v_1) = c(w)$. Necessarily, $c(w_1) = c(w_2)$; otherwise, we can recolor w and get case 6.3. So, we color u with $c(u) \in L(u) \setminus \{c(w), c(w_1)\}$ and v with $c(v) \in L(v) \setminus \{c(w), c(w_1), c(u)\}$.

The coloring obtained is acyclic.

7. Suppose that H contains the subgraph G_0 depicted in Figure 1. Let c be an acyclic L-coloring of $H \setminus \{u\}$. If $c(v) \neq c(u')$, we color then u with a color different from $c(v)$ and $c(u')$ and the coloring obtained is acyclic. So, $c(v) = c(u')$. If we cannot color u, this implies that w.l.o.g. $L(u) = \{1, 2, 3, 4\}$, $c(u') = c(v) = c(t) = 1$, $c(z_1) = 2$, $c(z_2) = 3$, and $c(s) = 4$. Observe that $L(v) = \{1, 2, 3, 4\}$; otherwise, we recolor v with a color different from 1, 2, 3, 4, and we color u with a color different from $c(v)$ and $c(u')$. The vertex r is colored with a color different from 1 and 4. Now, we recolor v with the color 4; we recolor s with a color different from $c(r), c(t), 4$, and we color u with a color different from $c(v)$ and $c(u')$. The coloring obtained is acyclic.

8. Suppose that H contains the subgraph G_1 depicted in Figure 2. Let c be an acyclic L-coloring of $H \setminus \{u\}$. We consider the following cases:

8.1. The colors of u_1, v, w are pairwise distinct. We color u with a color different from $c(u_1), c(v), c(w)$.

8.2. Suppose that $c(u_1) = c(w)$. If we cannot color u, this implies that w.l.o.g. $L(u) = \{1, 2, 3, 4\}$, $c(v) = 2$, $c(u_1) = c(w) = c(x_1) = 1$, $c(x) = 3$, and $c(y) = 4$. We erase the color of x. We recolor w with a color different from 1, 2, 4, then we recolor x with a color different from $c(w), c(y), c(x_1)$ and we color u with a color different from $c(u_1), c(v), c(w)$.

8.3. Suppose that $c(u_1) = c(v)$. If we cannot color u, this implies that w.l.o.g. $L(u) = \{1, 2, 3, 4\}$, $c(v) = c(u_1) = 1$, $c(w) = 2$, $c(v_1) = 3$, and $c(v_2) = 4$ ($c(x) \neq 2$, $c(y) \neq 2$). Observe that $L(v) = \{1, 2, 3, 4\}$; otherwise, we recolor v with a color different from $c(u_1), c(v_1), c(v_2), c(w)$ and we color u with a color different from $c(u_1), c(v), c(w)$. So we erase the color of w and we color v with 2. We then recolor w with a color different from $c(v), c(x), c(y)$. If $c(w) \neq 1$, we color u with a color different from $c(u_1), c(v), c(w)$, and the coloring obtained is acyclic. So, $c(w) = 1$ and consequently if we cannot color u, we have w.l.o.g. $c(x) = 3$, $c(y) = 4$, $c(x_1) = 1$, and we get case 8.2.

The extended coloring is acyclic.

9. Suppose that H contains the subgraph G_2 depicted in Figure 3. Let c be an acyclic L-coloring of $H \setminus \{u\}$. We consider the following cases:

9.1. The colors of u_1, v, w are pairwise distinct. We color u with a color different from $c(u_1), c(v), c(w)$.

9.2. Suppose that $c(u_1) = c(w)$. If we cannot color u, then w.l.o.g. $L(u) = \{1, 2, 3, 4\}$, $c(u_1) = c(w) = c(t_1) = 1$, $c(v) = 2$, $c(y) = 3$, and $c(t) = 4$. We know that one of the vertices x, z is colored by 1; if not, we can color u with 3. Moreover, $L(w) = \{1, 2, 3, 4\}$ (if not, we can recolor w and color u). Observe now that $L(v) = \{1, 2, 3, 4\}$, $c(v_1) = 3$ and $c(v_2) = 4$; otherwise we recolor w with 2 and v with a color different from $c(v_1), c(v_2), c(w)$ (remark that by the initial coloring, $c(v_1) \neq 2$ and $c(v_2) \neq 2$); if $c(v)$ is different from 1 and 2, we color u with a color different from $1, 2, c(v)$ and the coloring obtained is acyclic; hence $c(v) = 2$, $c(v_1) = 3$ and $c(v_2) = 4$.

9.2.1. Suppose that $c(x) = 1$ and so $c(x_1) = 3$. The vertex z is colored with a, $a \neq 1$, $a \neq 3$. If $a = 4$, then we recolor x with a color different from 1, 3, 4 and we color u with 3. So, $a \neq 1, 3, 4$. Observe that $L(y) = \{1, 3, 4, a\}$; otherwise we recolor y with a color different from $1, 3, 4, a$ and we color u with 3. Hence, we recolor y with 1, w with 3, x with a color different from 1, 3, a, and we color u with 4. The extended coloring is acyclic.

9.2.2. Suppose that $c(z) = 1$ and $c(z_1) = 3$ ($c(z_2) \neq 1, 3$).

9.2.2.1. Suppose that $c(x_1) = 1$. First, we erase the color of x. We recolor z with a color different from 1, 3, $c(z_2)$. If $c(z) \neq 4$, then we can extend the coloring to x and u: we color x with a color different from 1, 3, $c(z)$, and u with 3. So, $c(z) = 4$. If

$c(t_2) \neq 3$, we can color u and x as before. Now, $c(t_2) = 3$. In this case, we recolor t with a color different from 1, 3, 4 and we color u with 3 and x with a color different from 1, 3, 4.

9.2.2.2. Suppose that $c(x_1) \neq 1$. First, we erase the color of x. We recolor y with a color different from 1, 3, 4. If $c(y) \neq c(x_1)$, then we color u with 3 and x with a color different from 1, $c(y), c(x_1)$. The coloring obtained is acyclic. So, $c(y) = c(x_1)$. In this case, we color u with 3 and x with a color different from 1, 4, $c(y)$.

9.3. Suppose that $c(u_1) = c(v)$. If we cannot color u, then w.l.o.g. $L(u) = \{1,2,3,4\}$, $c(u_1) = c(v) = 1$, $c(w) = 2$, $c(v_1) = 3$ and $c(v_2) = 4$ ($c(y) \neq 2$ and $c(t) \neq 2$). Observe that $L(v) = \{1,2,3,4\}$; otherwise we recolor v with a color different from 1, 2, 3, 4, and we color u with a color different from 1, 2, $c(v)$. Now, we erase the color of w. We color v with 2 and w with a color different from $2, c(t), c(y)$. Observe that necessarily $c(w) = 1$, $\{c(y), c(t)\} = \{3,4\}$ since, otherwise, we can color u. Thus get case 9.2.

10. Suppose that H contains the subgraph G_3 depicted in Figure 4. Let c be an acyclic L-coloring of $H \setminus \{u\}$. If $c(u_1) \neq c(v)$, we color u with $c(u) \in L(u) \setminus \{c(u_1), c(v)\}$. Hence, $c(u_1) = c(v)$. Now, if we cannot color u, this implies that w.l.o.g. $L(u) = \{1,2,3,4\}$, $c(u_1) = c(v) = c(x_1) = 1$, $c(x) = 2$, $c(w) = 3$, $c(v_1) = 4$. Observe that necessarily, $L(v) = \{1,2,3,4\}$; otherwise we can recolor v with a color different from 1, 2, 3, 4, and we can color u with a color different from 1 and $c(v)$. So, we recolor v with 2, x with a color different from 1, 2, 3 and we color u with a color different from 1 and 2.

11. Suppose that H contains the subgraph G_4 depicted in Figure 5. Let c be an acyclic L-coloring of $H \setminus \{u, v, w\}$.

11.1. Assume that $c(y_1) \neq c(x)$. We erase the color of y. We color w with a color from its list, different from $c(x)$, which appears at most once on y_1, u_1, u_2, v_1, v_2. We color then the vertices u, v, y: we color each vertex with a color from its list different from the colors which appear in the neighbourhood and different from the color of x if necessary.

11.2. Assume that $c(y_1) = c(x)$. First, observe that $c(x_1) = c(x_2)$; otherwise, we can erase the color of y, recolor x with a color different from $c(x_1), c(x_2), c(y_1)$ and we get case 11.1. We erase the color of y. We color w with a color different from $c(x)$ which appears at most once on the vertices u_1, u_2, v_1, v_2. We color y with a color different from $c(y_1), c(w), c(x_1)$. Finally we color u and v with a color from their list different from the colors which appear in their neighbourhood and different from $c(x)$ if necessary.

The coloring obtained is acyclic.

12. Suppose that H contains the subgraph G_5 depicted in Figure 6. Let c be an acyclic L-coloring of $H \setminus \{u, v, w, z\}$.

12.1. Assume that $c(y_1) \neq c(x)$. We erase the color of y. We color z with a color different from $c(x)$ which appears at most once on the vertices u_1, v_1, w_1, w_2, y_1. We color then the vertices u, v, w, y with a color different from the colors of the neighbourhood and different from $c(x)$ if necessary.

12.2. Assume that $c(y_1) = c(x)$. If it exists, we color z with a color different from $c(x), c(y)$ which appears at most once on u_1, v_1, w_1, w_2; then we color the vertices u, v, w with a color different from the colors of the neighbourhood and different from $c(x)$ if necessary. So, if this color does not exist, this implies that w.l.o.g. $L(z) = \{1, 2, 3, 4\}$, $c(x) = c(y_1) = 1$, $c(y) = 2$, $c(u_1) = c(w_1) = 3$ and $c(v_1) = c(w_2) = 4$. In this case, we color z with 3, w with a color different from 1, 3, 4; we color v with a color different from 3, 4, and u with a color different from 1,3, $c(w)$.

The coloring obtained is acyclic.

13. Suppose that H contains a 5-vertex v adjacent to four 2-vertices u_1, u_2, u_3, u_4 and another vertex z. Each 2-vertex u_i is adjacent to v and u'_i, with $1 \leq i \leq 4$. Let c be an acyclic L-coloring of $H \setminus \{v, u_1, u_2, u_3, u_4\}$. We color v with a color different from $c(z)$ which appears at most once on u'_1, u'_2, u'_3, u'_4. Now, for $i = 1$ to 4, we color u_i with a color different from $c(u'_i)$ and $c(v)$ if $c(u'_i) \neq c(v)$ and with a color different from $c(u'_i)$ and $c(z)$ otherwise.

14. Suppose that H contains a 5-vertex v adjacent to three 2-vertices u_1, u_2, u_3 and two vertices z_1, z_2 such that $[vz_1z_2]$ is a 3-face. Each 2-vertex u_i is adjacent to v and u'_i for $1 \leq i \leq 3$. Let c be an acyclic L-coloring of $H \setminus \{v, u_1, u_2, u_3\}$. We color v with a color different from $c(z_1), c(z_2)$ which appears at most once on u'_1, u'_2, u'_3. Now, for $i = 1$ to 3, we color u_i with a color different from $c(u'_i)$ and $c(v)$ if $c(u'_i) \neq c(v)$ and with a color different from $c(u'_i), c(z_1)$ and $c(z_2)$ otherwise.

Lemma 2.2. *Let H be a connected plane graph with n vertices, m edges and r faces. Then we have the following:*

$$\sum_{v \in V(H)} (5d(v) - 14) + \sum_{f \in F(H)} (2r(f) - 14) = -28 \quad (1)$$

$$\sum_{v \in V(H)} (d(v) - 4) + \sum_{f \in F(H)} (r(f) - 4) = -8 \quad (2)$$

$$\sum_{v \in V(H)} (2d(v) - 6) + \sum_{f \in F(H)} (r(f) - 6) = -12 \quad (3)$$

Proof. Euler's formula $n - m + r = 2$ can be rewritten as $(10m - 14n) + (4m - 14r) = -28$, $(2m - 4n) + (2m - 4r) = -8$, and $(4m - 6n) + (2m - 6r) = -12$, respectively. These identities and the relation $\sum_{v \in V} d(v) = \sum_{f \in F} r(f) = 2m$ imply (1), (2), and (3), respectively.

3 Proof of Theorem 1.7

Let H be a counterexample to Theorem 1.7 of the minimum order. Let L be a list assignment such that $|L(v)| = 4$ for all $v \in V(H)$. Then H satisfies Lemma 2.1.

We define the weight function ω by $\omega(x) = 5d(x) - 14$ if $x \in V(H)$ and $\omega(x) = 2r(x) - 14$ if $x \in F(H)$. It follows from identity (1) that the total sum of weights is equal to -28. In the following, we will define discharging rules (R1) to (R4) and redistribute weights accordingly. Once the discharging is finished, a new weight function ω^* is produced. However, the total sum of weights is kept fixed when the discharging is in process. Nevertheless, we can show that $\omega^*(x) \geq 0$ for all $x \in V(H) \cup F(H)$. This leads to the following obvious contradiction

$$0 \leq \sum_{x \in V(H) \cup F(H)} \omega^*(x) = \sum_{x \in V(H) \cup F(H)} \omega(x) = -28 < 0$$

and hence demonstrates that no such counterexample can exist.

The discharging rules are defined as follows:

(R1) Every vertex v of degree at least 5 gives 4.5 to each incident 3-face, 2 to each adjacent 2-vertex, and 1 to each adjacent pendant light 3-vertex.

(R2) Let v be a 4-vertex.
If $m_3(v) = 2$, then v gives 3 to each incident 3-face.
If $m_3(v) = 1$, then v gives 4 to the unique incident 3-face, 2 to each adjacent 2-vertex, and 1 to each adjacent pendant light 3-vertex.
If $m_3(v) = 0$, then v gives 2 to the adjacent 2-vertex, and 1 to each adjacent pendant light 3-vertex.

(R3) Let v be a 3-vertex. If v is light, then v gives 2 to the incident 3-face. Otherwise, v gives 1 to the incident 3-face.

(R4) Each 3-face that is not light gives 0.5 to each of the opposite light 3-faces.

In order to complete the proof, it suffices to inspect that the new weight function $\omega^*(x)$ is non-negative for all $x \in V(H) \cup F(H)$. In the sequel, we use $o(f)$ to denote the number of opposite light 3-faces of a 3-face f. Let $p_3(v)$ denote the number of pendant light 3-vertices of a vertex v. It is an easy observation that $p_3(v) \leq n_3(v)$.

Let $v \in V(G)$ be a k-vertex. Then $k \geq 2$ by (C1).

- If $k = 2$, then $\omega(v) = -4$ and v is adjacent to two vertices of degree at least 4 by (C3). By (R1), (R2), and (C2), $\omega^*(v) = -4 + 2 + 2 = 0$.
- If $k = 3$, then $\omega(v) = 1$. If v is not light, then $\omega^*(v) \geq 1 - 1 = 0$ by (R3). Suppose that v is light. There is a 3-face $[vuy]$ such that $d(u) = 3$ and $d(y) \geq 5$. Let x be the neighbour of v in G that differs from u and y. By (C2) to (C4), we derive that $d(x) \geq 4$. Note that the edge vx can not be incident to any 3-face for, otherwise, G will contain a 4-cycle. This implies that v is a pendant light 3-vertex of x. By (R1) and (R2), x sends 1 to v, and hence $\omega^*(v) \geq 1 + 1 - 2 = 0$ by (R3).

- Assume that $k = 4$. We see that $\omega(v) = 6$ and $m_3(v) \le 2$. Moreover, v is adjacent to at most one 2-vertex by (C5). If v is adjacent to a 2-vertex, then v is not adjacent to any pendant light 3-vertex by (C7). So it follows from these facts and (R2) that $\omega^*(v) \ge 6 - 2 - 1 \times 3 = 1$ if $m_3(v) = 0$, $\omega^*(v) \ge 6 - 4 - 2 = 0$ if $m_3(v) = 1$, and $\omega^*(v) \ge 6 - 3 - 3 = 0$ if $m_3(v) = 2$.
- Assume that $k \ge 5$. On the one hand, we note that $m_3(v) \le \lfloor k/2 \rfloor$ since G contains no two adjacent 3-faces. On the other hand, there is no a 2-vertex or a pendant light 3-vertex on the boundary of any 3-face incident to v. This implies that $n_2(v) + p_3(v) \le k - 2m_3(v)$. Since, by (R1),

$$\begin{aligned} & 4.5m_3(v) + 2n_2(v) + p_3(v) \\ \le\ & 4.5m_3(v) + 2(n_2(v) + p_3(v)) \\ \le\ & 4.5m_3(v) + 2(k - 2m_3(v)) \\ \le\ & 2k + 0.5m_3(v) \le 2k + 0.5\lfloor k/2 \rfloor \\ \le\ & 5k - 14, \end{aligned}$$

  it follows that $\omega^*(v) = \omega(v) - 4.5m_3(v) + 2n_2(v) + p_3(v) \ge 0$.

Suppose that $f \in F(H)$ is a k-face. Then $k \ne 4, 5, 6$.

- If $k \ge 7$, then $\omega^*(f) = \omega(f) = 2k - 14 \ge 0$.
- Assume that $k = 3$. Then $\omega(f) = -8$. We write $f = [xyz]$ such that $d(x) \le d(y) \le d(z)$. By (C2), $d(x) \ge 3$. If $d(y) = 3$, then $d(z) \ge 5$ by (C6) and both x and y are light 3-vertices. By definition, f has no an opposite light 3-face, i.e., $o(f) = 0$. We get $\omega^*(f) \ge -8 + 4.5 + 2 + 2 = 0.5$ by (R1) and (R3). If $d(y) \ge 5$, then since $o(f) \le 1$ and each of y and z sends 4.5 to f, we have $\omega^*(f) \ge -8 + 1 + 2 \times 4.5 - 0.5 = 1.5$ by (R1) to (R4). Thus, from now on, we suppose that $d(y) = 4$. The proof is divided into the following subcases:
  1. Assume that $d(x) = 3$ and $d(z) \ge 5$. Since at most y admits an opposite light 3-face, we have $\omega^*(f) \ge -8 + 4.5 + 3 + 1 - 0.5 = 0$.
  2. Assume that $d(x) = 4$ and $d(z) \ge 5$. We see that $o(f) \le 2$ and henceforth $\omega^*(f) \ge -8 + 4.5 + 2 \times 3 - 2 \times 0.5 = 1.5$.
  3. Assume that $d(x) = 3$ and $d(z) = 4$. If $m_3(y) = m_3(z) = 1$, then $o(f) = 0$ and $\omega^*(f) \ge -8 + 2 \times 4 + 1 = 1$. If $m_3(y) = 1$ and $m_3(z) = 2$, then (C8) asserts that $o(f) = 0$, and therefore $\omega^*(f) \ge -8 + 4 + 3 + 1 = 0$. If $m_3(y) = 2$ and $m_3(z) = 1$, we have a similar proof. If $m_3(y) = m_3(z) = 2$, i.e., f is a light 3-face, then f receives 0.5 from each of its two opposite 3-faces. Consequently, $\omega^*(f) \ge -8 + 2 \times 3 + 1 + 0.5 \times 2 = 0$ by (R2) to (R4).
  4. Finally, assume that $d(x) = d(z) = 4$. (C9) asserts that $o(f) \le 2$ and hence $\omega^*(f) \ge -8 + 3 \times 3 - 2 \times 0.5 = 0$.

4 Proof of Theorem 1.8

Let H be a counterexample to Theorem 1.8 of the minimum order. Let L be a list assignment such that $|L(v)| = 4$ for all $v \in V(H)$. Then H satisfies Lemma 2.1.

We apply the formula (2) in Lemma 2.2 and thus define the weight function ω by $\omega(x) = d(x) - 4$ for all $x \in V(H)$ and $\omega(x) = r(x) - 4$ for all $x \in F(H)$. New discharging rules are defined as follows:

(R1) Every vertex v of degree at least 5 gives $\frac{1}{3}$ to each adjacent 2-vertex, and $(\omega(v) - \frac{1}{3}n_2(v))/n_3(v)$ to each adjacent 3-vertex.

Let $\beta(v)$ denote the total weight of a 2-vertex or a 3-vertex v obtained from its adjacent vertices of degree at least 5 according to (R1).

(R2) Every face f of degree 6 or at least 8 gives $\frac{1}{3}$ to each adjacent 3-face, $\frac{1}{2}(2 - \beta(v))$ to each incident 2-vertex v, and $(1 - \beta(u))/(3 - m_3(u))$ to each incident 3-vertex u provided $\beta(u) < 1$.

Observation 4.1. *(R1) and (R2) possess the following properties:*

(O1) *Every vertex v of degree at least 6 gives at least $\frac{1}{3}$ to each adjacent 3-vertex (since $\omega(v) - \frac{1}{3}(n_2(v) + n_3(v)) \geq d(v) - 4 - \frac{1}{3}d(v) \geq 0$ whenever $d(v) \geq 6$).*
(O2) *Every face f of degree 6 or at least 8 gives at most 1 to each incident 2-vertex, at most $\frac{1}{2}$ to each incident 3-vertex u with $m_3(u) = 1$, and at most $\frac{1}{3}$ to each incident 3-vertex u with $m_3(u) = 0$.*

The following Observation 4.2 gives some useful facts, which are either straightforward or the immediate consequences of the previous arguments.

Observation 4.2. *Suppose that f is a k-face with $k \geq 6$. Then*

(O3) $n_2(f) \leq \lfloor k/2 \rfloor$;
(O4) $n_3(f) \leq \lfloor 2k/3 \rfloor$;
(O5) $n_3(f) + 2n_2(f) + 1 \leq k$;
(O6) $m_3(f) \leq \lfloor k - 2n_2(f) - \frac{1}{2}n_3(f) \rfloor$.

We first carry out (R1), and then (R2) in the graph G. It suffices to show that the resultant weight function $\omega^*(x)$ is non-negative for all $x \in V(G) \cup F(G)$.

Let v be a k-vertex with $k \geq 2$:

- If $k = 2$, then $\omega(v) = -2$. Since v is incident to two faces of degree at least 6 by (C2), it holds clearly that $\omega^*(v) = 0$ by (R1) and (R2).
- If $k = 3$, then $\omega(v) = -1$. Noting that v is incident to at most one 3-face, i.e., $m_3(v) \leq 1$, we have that $\omega^*(v) = 0$ by (R1) and (R2).
- If $k = 4$, then $\omega^*(v) = \omega(v) = 0$.

- If $k = 5$, then $\omega(v) = 1$. Since v is adjacent to at most three 2-vertices by (C13), it follows from (R1) that $\omega(v) - \frac{1}{3}n_2(v) \geq 1 - 3 \times \frac{1}{3} = 0$, which implies that $\omega^*(v) \geq 0$.
- If $k \geq 6$, then $\omega(v) - \frac{1}{3}n_2(v) \geq \omega(v) - \frac{1}{3}k = \frac{2}{3}k - 4 \geq 0$ and thus $\omega^*(v) \geq 0$ by (R1).

Let f be a k-face with $k \geq 3$. Then $k \neq 4, 5, 7$. We write $f = [x_1 x_2 \cdots x_k]$, where $x_1, x_2, \cdots, x_k$ are the boundary vertices of f in the cyclic order.

- If $k = 3$, then $\omega(f) = -1$. Each of the adjacent faces of f is of degree at least 8 for G contains no 4-cycles and 7-cycles. Thus each face adjacent to f gives $\frac{1}{3}$ to f by (R2) so that $\omega^*(f) = -1 + 3 \times \frac{1}{3} = 0$.
- Suppose that $k = 6$. We see that $\omega(f) = 2$, and $n_2(f) \leq 3$ by (O3). Since f is not adjacent to any 3-face, there is no a 3-vertex on the boundary of f which is incident to a 3-face. First suppose that $n_2(f) = 3$, w.l.o.g., $d(x_i) = 2$ for $i = 1, 3, 5$. By (C3) and (C5), $d(x_j) \geq 5$ for $j = 2, 4, 6$. Thus $\beta(x_i) = \frac{2}{3}$ by (R1) and f only needs to give $\frac{2}{3}$ to x_i for $i = 1, 3, 5$. It turns out that $\omega^*(v) = 2 - 3 \times \frac{2}{3} = 0$ by (R2). Next suppose that $n_2(f) = 2$. It is easy to show that f is incident to at most one 3-vertex by (C2). When $n_3(f) = 0$, $\omega^*(f) \geq 2 - 2 \times 1 = 0$. When $n_3(f) = 1$, there exists some index i such that $d(x_{i-1}) = d(x_{i+1}) = 2$ and $d(x_i) \geq 5$, where indices are taken modulo 6. So x_i gives $\frac{1}{3}$ to each of x_{i-1} and x_{i+1} by (R1), and furthermore f gives at most $\frac{5}{6}$ to each of them. Consequently, $\omega^*(f) \geq 2 - 2 \times \frac{5}{6} - \frac{1}{3} = 0$. Now suppose that $n_2(f) = 1$. We see $n_3(f) \leq 2$ by (C3) and (C4), and thus $\omega^*(f) \geq 2 - 1 - 2 \times \frac{1}{3} = \frac{1}{3}$ by the previous property (O2). Finally suppose that $n_2(f) = 0$. (C4) asserts that $n_3(f) \leq 4$ and henceforth $\omega^*(f) \geq 2 - 4 \times \frac{1}{3} = \frac{2}{3}$.
- Suppose that $k = 8$. Then $\omega(f) = 4$, $n_2(f) \leq 4$, and $n_3(f) \leq 5$. If $n_2(f) = 4$, then $n_3(f) = m_3(f) = 0$ and $\omega^*(f) \geq 4 - 4 = 0$. Assume that $n_2(f) = 3$. It follows that $n_3(f) \leq 1$ and $m_3(f) \leq 2$. When $n_3(f) = 0$, we have $\omega^*(f) \geq 4 - 3 - 2 \times \frac{1}{3} = \frac{1}{3}$. When $n_3(f) = 1$, it is easy to check that $m_3(f) \leq 1$ and hence $\omega^*(f) \geq 4 - 3 - \frac{1}{2} - \frac{1}{3} = \frac{1}{6}$. Assume that $n_2(f) = 2$. Note that $n_3(f) \leq 2$ by (C4), and $m_3(f) \leq 4$. If $n_3(f) \leq 1$, then $\omega^*(f) \geq 4 - 2 - \frac{1}{2} - 4 \times \frac{1}{3} = \frac{1}{6}$. If $n_3(f) = 2$, then $m_3(f) \leq 3$ and thus $\omega^*(f) \geq 4 - 2 - 2 \times \frac{1}{2} - 3 \times \frac{1}{3} = 0$. Assume that $n_2(f) = 1$. It is easy to see that $n_3(f) \leq 4$ and $m_3(f) \leq 6$. Without loss of generality, we suppose that $d(x_1) = 2$. So $d(x_2) \geq 4$ and $d(x_8) \geq 4$ by (C3). If $n_3(f) \leq 2$, then $\omega^*(f) \geq 4 - 1 - 2 \times \frac{1}{2} - 6 \times \frac{1}{3} = 0$. If $n_3(f) = 3$, then $m_3(f) \leq 4$ and $\omega^*(f) \geq 4 - 1 - 3 \times \frac{1}{2} - 4 \times \frac{1}{3} = \frac{1}{6}$. If $n_3(f) = 4$, then $m_3(f) \leq 4$ and the only possible case is that $d(x_3) = d(x_4) = d(x_6) = d(x_7) = 3$ and $d(x_5) \geq 4$. Observe that if $m_3(f) \leq 3$, then $\omega^*(f) \geq 0$. If $m_3(f) = 4$, by (C10), $d(x_2) \geq 5$ and $d(x_8) \geq 5$. (R1) asserts that each of x_2 and x_8 gives $\frac{1}{3}$ to x_1, which implies that f gives x_1 at most $\frac{2}{3}$. Thus $\omega^*(f) \geq 4 - \frac{2}{3} - 4 \times \frac{1}{2} - 4 \times \frac{1}{3} = 0$. Finally we assume that $n_2(f) = 0$. It is easy to derive that $n_3(f) \leq 5$ and $m_3(f)$ is at most 8. If $n_3(f) \leq 2$, then $\omega^*(f) \geq 4 - 2 \times \frac{1}{2} - 8 \times \frac{1}{3} = \frac{1}{3}$. If $3 \leq n_3(f) \leq 4$, then $m_3(f) \leq 6$ and $\omega^*(f) \geq 4 - 4 \times \frac{1}{2} - 6 \times \frac{1}{3} = 0$.

Suppose now that $n_3(f) = 5$. Then $m_3(f) \le 5$ in this case. If $m_3(f) \le 4$, then $\omega^*(f) \ge 4 - \frac{5}{2} - \frac{4}{3} = \frac{1}{6}$. So we assume that $m_3(f) = 5$ in the following discussion. In view of the structure of f, we may assume, w.l.o.g., that $d(x_i) = 3$ for all $i = 1,3,4,6,7$, $d(x_j) \ge 4$ for all $j = 2,5,8$, and f is adjacent to two 3-faces $[x_4x_5y]$ and $[x_5x_6z]$. It follows from (C2) and (C4) that $d(y) \ge 4$ and $d(z) \ge 4$. If $d(x_5) \ge 5$, then x_5 gives at least $\frac{1}{3}$ to each of x_4 and x_6 by (R1). Furthermore, f gives each of them at most $\frac{1}{3}$ by (R2). Consequently, $\omega^*(f) \ge 4 - \frac{3}{2} - \frac{2}{3} - \frac{5}{3} = \frac{1}{6}$. So suppose that $d(x_5) = 4$. By (C8), both $d(y)$ and $d(z)$ are at least 5. Since y cannot be adjacent to three 2-vertices by (C14), y gives x_4 at least $\frac{1}{6}$. Similarly, z gives x_6 at least $\frac{1}{6}$. Consequently, f gives at most $\frac{5}{12}$ to each of x_4 and x_6, therefore $\omega^*(f) \ge 4 - \frac{5}{3} - \frac{3}{2} - 2 \times \frac{5}{12} = 0$.

- Suppose that $k = 9$. It is easy to see that $\omega(f) = 5$, $n_2(f) \le 4$, and $n_3(f) \le 6$. If $n_2(f) = 4$, then $n_3(f) = 0$ by (O5), and $m_3(f) \le 1$ by (O6). Thus $\omega^*(f) \ge 5 - 4 - \frac{1}{3} = \frac{2}{3}$. If $n_2(f) = 3$, then $n_3(f) \le 2$ and $m_3(f) \le 3$. Thus $\omega^*(f) \ge 5 - 3 - 2 \times \frac{1}{2} - 3 \times \frac{1}{3} = 0$. Assume that $n_2(f) = 2$. It is immediate to derive that $n_3(f) \le 4$ and $m_3(f) \le 5$. If $n_3(f) \le 2$, then $\omega^*(f) \ge 5 - 2 - 2 \times \frac{1}{2} - 5 \times \frac{1}{3} = \frac{1}{3}$. If $3 \le n_3(f) \le 4$, then it is easy to check that $m_3(f) \le 3$ and hence $\omega^*(f) \ge 5 - 2 - 4 \times \frac{1}{2} - 3 \times \frac{1}{3} = 0$. Assume that $n_2(f) = 1$. Then $n_3(f) \le 4$ by (C4), and $m_3(f) \le 7$ by (O6). If $m_3(f) = 7$, then it is easily seen that $n_3(f) = 0$ and thus $\omega^*(f) \ge 5 - 1 - 7 \times \frac{1}{3} = \frac{5}{3}$; otherwise, $\omega^*(f) \ge 5 - 1 - 4 \times \frac{1}{2} - 6 \times \frac{1}{3} = 0$. Assume that $n_2(f) = 0$. We see that $n_3(f) \le 6$ in this case. If $n_3(f) \le 4$, then $\omega^*(f) \ge 5 - 4 \times \frac{1}{2} - 9 \times \frac{1}{3} = 0$. If $5 \le n_3(f) \le 6$, then $m_3(f) \le 6$, and therefore $\omega^*(f) \ge 5 - 6 \times \frac{1}{2} - 6 \times \frac{1}{3} = 0$.
- Suppose that $k \ge 10$. Using (O2) to (O6), we can establish the following estimate:

$$
\begin{aligned}
& n_2(f) + \frac{1}{2}n_3(f) + \frac{1}{3}m_3(f) \\
\le\ & n_2(f) + \frac{1}{2}n_3(f) + \frac{1}{3}(k - 2n_2(f) - \frac{1}{2}n_3(f)) \\
=\ & \frac{1}{3}k + \frac{1}{6}(2n_2(f) + n_3(f)) + \frac{1}{6}n_3(f) \\
\le\ & \frac{1}{3}k + \frac{1}{6}(k-1) + \frac{1}{6}\lfloor 2k/3 \rfloor \\
\le\ & \frac{11}{18}k - \frac{1}{6} \le k - 4.
\end{aligned}
$$

Thus we always have that $\omega^*(f) \ge \omega(f) - (n_2(f) + \frac{1}{2}n_3(f) + \frac{1}{3}m_3(f)) \ge 0$.

Up to now, we have proved that, for all $x \in V(H) \cup F(H)$, $\omega^*(x) \ge 0$. Therefore, by Equation (2) in Lemma 2.2, we drive the following obvious contradiction, which completes the proof of the theorem.

$$0 \le \sum_{x \in V(H) \cup F(H)} \omega^*(x) = \sum_{x \in V(H) \cup F(H)} \omega(x) = -8.$$

5 Proof of Theorem 1.9

Let H be a counterexample to Theorem 1.9 of the minimum order. Let L be a list assignment such that $|L(v)| = 4$ for all $v \in V(H)$. Then H satisfies Lemma 2.1.

A *light 5-vertex* of G is a 5-vertex incident to a 3-face and adjacent to two 2-vertices and one pendant light 3-vertex. A *weak 3-face* is a 3-face incident to a 3-vertex.

This time we define the weight function ω by $\omega(x) = 2d(x) - 6$ if $x \in V(H)$ and $\omega(x) = r(x) - 6$ if $x \in F(H)$. It follows from identity (3) in Lemma 2.2 that the total sum of weights is equal to -12. We then apply the following discharging rules:

(R0) Every vertex of degree at least 6 gives 2 to each incident 3-face, 1 to each adjacent 2-vertex, and 0.5 to each adjacent pendant light 3-vertex.

(R1) Let v be a 5-vertex.

If v is not light, then v gives 2 to each incident 3-face, 1 to each adjacent 2-vertex, and 0.5 to each adjacent pendant light 3-vertex.

If v is light, then v gives 1 to each incident 3-face, 1 to each adjacent 2-vertex, and 0.5 to each adjacent pendant light 3-vertex.

(R2) Let v be a 4-vertex.

If $m_3(v) = 0$, then v gives 1 to each adjacent 2-vertex and 0.5 to each adjacent pendant light 3-vertex.

Assume that $m_3(v) = 1$ and the incident 3-face is weak. If v is adjacent to two pendant light 3-vertices, then v gives 1 to the incident 3-face and 0.5 to each pendant light 3-vertex. If v is adjacent to at most one pendant light 3-vertex, then v gives 1.5 to the incident 3-face and 0.5 to the pendant light 3-vertex

Assume that $m_3(v) = 1$ and the incident 3-face is not weak. The vertex v gives 1 to the incident 3-face, 1 to each adjacent 2-vertex and 0.5 to each adjacent pendant light 3-vertex.

(R3) Each light 3-vertex gives 0.5 to the incident 3-face.

Let $\omega^*(x)$ denote the new weight function once the discharging procedure is complete. To complete the proof of the theorem, it remains to check that the resultant weight function $\omega^*(x)$ is non-negative for all $x \in V(G) \cup F(G)$.

Let v be a k-vertex with $k \geq 2$.

- If $k = 2$, then $\omega(v) = -2$ and v receives 1 from each neighbour since v is adjacent to two vertices of degree at least 4 by (C3), and we note that if a 4-vertex is incident to a weak 3-face, then it is not adjacent to a 2-vertex by (C10). Thus, $\omega^*(v) = -2 + 2 \cdot 1 = 0$ by (R0) to (R2).
- If $k = 3$, then $\omega(v) = 0$. If v is not light, then $\omega^*(v) = \omega(v) = 0$. Now, if v is light, then it receives at least 0.5 from its neighbourhood (since v is not adjacent to a 2-vertex by (C3) and v is adjacent to at most one 3-vertex by (C4)) and gives 0.5 to the incident 3-face. So, $\omega^*(v) = 0$.

- If $k = 4$, then $\omega(v) = 2$. The vertex v is adjacent to at most one 2-vertex by (C5). Moreover, if v is adjacent to a 2-vertex, then v is incident to neither a weak 3-face by (C10) nor a light 3-vertex by (C7). If v is not incident to a 3-face, then it sends out at most 2 and $\omega^*(v) \geq 0$. If v is incident to a 3-face, which is not weak, then it sends out at most 2 and $\omega^*(v) \geq 0$ (it gives 1 to the incident 3-face and it gives 1 to the adjacent 2-vertex or at most 2×0.5 to the pendant light 3-vertices). If v is incident to a weak 3-face (and so v is not adjacent to a 2-vertex), v sends out at most 2 and $\omega^*(v) \geq 0$ by (R2).
- If $k = 5$, then $\omega(v) = 4$. If v is light, then it sends out 3.5 and so $\omega^*(v) = 0.5$. Assume that v is not light. The vertex v is adjacent to at most three 2-vertices by (C13); moreover, if v is adjacent to three 2-vertices, then it is not incident to a 3-face by (C14). Consequently, v sends out at most 4 by (R1) and $\omega^*(v) \geq 0$.
- If $k \geq 6$, then it is easy to see that v sends out at most k by (R0). Since $k \leq 2k - 6 = \omega(v)$, we have $\omega^*(v) \geq 0$.

Let f be a k-face. Then $k \neq 4, 5$. If $k \geq 6$, then $\omega(f) = \omega^*(f) \geq 0$. Suppose that $k = 3$. Let $f = [xyz]$ denote this face such that $d(x) \leq d(y) \leq d(z)$. By (C2), $d(x) \geq 3$. Now, if $d(x) = 3$ and $d(y) = 3$, by (C4) and (C6), $d(z) \geq 5$. We have to consider the following cases:

- $d(x) = 3, d(y) = 3$, and $d(z) \geq 5$. The vertices x and y are light by definition and z cannot be a light 5-vertex by (C12). So f receives $2 \times 0.5 + 2 = 3$ and $\omega^*(f) = 0$.
- $d(x) = 3, d(y) = 4$, and $d(z) = 4$. The face f is a weak 3-face. By (C11), neither y nor z are adjacent each to two pendant light 3-vertices. So, y and z give each 1.5 to the 3-face and $\omega^*(f) = 0$.
- $d(x) = 3, d(y) = 4$, and $d(z) \geq 5$. The face f is a weak 3-face. By (C12), z is not a light 5-vertex. So, y gives at least 1 to f and z gives 2. Hence, $\omega^*(f) \geq 0$.
- $d(x) \geq 4, d(y) \geq 4$, and $d(z) \geq 4$. Each vertex gives at least 1 by (R0) to (R2) and $\omega^*(f) \geq 0$.

For all $x \in V(H) \cup F(H)$, $\omega^*(x) \geq 0$. The contradiction obtained with identity (3) in Lemma 2.2 completes the proof.

6 Concluding Remarks

An *oriented k-coloring* of an oriented graph $\vec{G}$ is a partition of vertex set into k color classes such that no two adjacent vertices belong to the same color class and all the arcs linking two color classes have the same direction. The *oriented chromatic number* of an oriented graph $\vec{G}$, denoted by $\chi_o(\vec{G})$, is the smallest k such that $\vec{G}$ has an oriented k-coloring. The *oriented chromatic number* $\chi_o(G)$ of an undirected graph G is defined as the maximum oriented chromatic number of its orientations.

Raspaud and Sopena [RS94] established an interesting relation between the oriented chromatic number and the acyclic chromatic number of a graph G: if $\chi_a(G) = k$, then $\chi_o(G) \leq k2^{k-1}$. Combining this fact with Theorem 1.1, they observed that the oriented chromatic number of a planar graph is at most 80. Since, for any graph G, $\chi_a(G) \leq \chi_a^l(G)$, our Theorems 1.7 to 1.9 imply clearly the following results concerning the oriented chromatic number of planar graphs.

Theorem 6.1. *Suppose that G is a planar graph without 4-cycles and 5-cycles. If moreover G does not contain one of the following configurations, then $\chi_o(G) \leq 32$:*

(1) *a 6-cycle;*
(2) *a 7-cycle;*
(3) *two intersecting triangles.*

It is unknown if Theorems 1.7 to 1.9 are the best possible in the sense that there exist planar graphs G satisfying the corresponding conditions about cycles such that $\chi_a^l(G) = 4$. However, we like to put forward the following conjecture:

Conjecture 6.2 (*"Domaine de la Solitude 2000"* 's Conjecture). *Every planar graph without 4-cycles is acyclically 4-choosable.*

Lam, Xu and Liu [LXL99] proved that every planar graph G without 4-cycles is 4-choosable. Conjecture 6.2, if it is true, will strengthen this result.

Another challenging problem regarding the acyclic choosability of planar graphs is as follows:

Problem 6.3. Is acyclically 5-choosable every planar triangle-free graph ?

As a special case of Problem 6.3, the following question seems more interesting:

Problem 6.4. Is acyclically 5-choosable every bipartite planar graph ?

In view of examples of bipartite planar graphs constructed in [KM76], the number 5 in Problems 6.3 and 6.4 is the best possible obviously.

References

[AB77] M. O. Albertson and D. M. Berman, Every planar graph has an acyclic 7-coloring, *Israel J. Math.*, (28):169–174, 1977.

[BFKRS02] O. V. Borodin, D. G. Fon-Der Flaass, A. V. Kostochka, A. Raspaud, and E. Sopena, Acyclic list 7-coloring of planar graphs, *J. Graph Theory*, 40(2):83–90, 2002.

[BKW99] O. V. Borodin, A. V. Kostochka, and D. R. Woodall, Acyclic colourings of planar graphs with large girth, *J. London Math. Soc.*, 2(60):344–352, 1999.

[Bor79] O. V. Borodin, On acyclic coloring of planar graphs, *Discrete Math.*, 25:211–236, 1979.

[Grü73] B. Grünbaum, Acyclic colorings of planar graphs, *Israel J. Math.*, 14:390–408, 1973.

[KM76] A. V. Kostochka and L. S. Mel'nikov, Note to the paper of Grünbaum on acyclic colorings, *Discrete Math.*, 14:403–406, 1976.

[Kos76] A. V. Kostochka, Acyclic 6-coloring of planar graphs, *Metody Diskret. Anal.*, (28):40–56, 1976. In Russian.

[LXL99] P. Lam, B. Xu, and J. Liu, The 4-choosability of planar graphs without 4-cycles, *J. Combin. Theory Ser. B*, 76:117–126, 1999.

[Mit74] J. Mitchem, Every planar graph has an acyclic 8-coloring, *Duke Math. J.*, 41(1):177–181, 1974.

[MOR05] M. Montassier, P. Ochem, and A. Raspaud, On the acyclic choosability of graphs, *Journal of Graph Theory*, 2005. To appear.

[RS94] A. Raspaud and E. Sopena, Good and semi-strong colorings of oriented planar graphs, *Inform. Process. Lett.*, 51:171–174, 1994.

On the Algorithmic Aspects of Hedetniemi's Conjecture

Claude Tardif

Royal Military College of Canada, PO Box 17000 Stn Forces, Kingston, ON, K7K 7B4 Canada
Claude.Tardif@rmc.ca

Summary. We present a polynomial algorithm, implicit in the work of El-Zahar and Sauer, which inputs a 3-colouring of a categorical product of two graphs and outputs a 3-colouring of one of the factors. We raise a question about the existence of polynomial algorithms for colouring the vertices of some graphs in terms of intrinsic succint description of the vertices rather than in terms of the (exponential) size of the graph.

AMS Subject Classification. 05C15.
Keywords. Graph colourings, graph products, Hedetniemi's conjecture.

1 Introduction

The *categorical product* $G \times H$ of two graphs G and H is the graph with vertex-set $V(G) \times V(H)$, where $[(u,v),(u',v')] \in E(G \times H)$ if and only if $[u,u'] \in E(G)$ and $[v,v'] \in E(H)$. Hedetniemi's conjecture [Hed66] states that

$$\chi(G \times H) = \min\{\chi(G), \chi(H)\}. \tag{1}$$

It is easy to see (1) holds whenever the right-hand term is at most 3, using the fact that a graph is bipartite if and only if it does not contain an odd cycle. However as no such criteria are available for larger chromatic numbers, it is not clear that (1) holds for larger values of $\min\{\chi(G), \chi(H)\}$. A considerable breakthrough was achieved by El-Zahar and Sauer [ES85] who proved that the chromatic number of the categorical product of two 4-chromatic graphs is 4. In 1999 Jack Edmonds pointed out to the author that the contrapositive formulation of this result raises an interesting algorithmic question: Given two graphs G and H and a 3-colouring $c : G \times H \to \{0,1,2\}$ of their product, how hard is it to find a 3-colouring of G or of H. The purpose of this note is to provide the following answer.

Proposition 1.1. *There exists a polynomial algorithm which imputs two graphs G and H and a 3-colouring $c : G \times H \to \{0,1,2\}$ of their product, and outputs a 3-colouring of G or of H.*

The simple algorithm is given in the next section, and the proof of its correctness is given in Section 3, drawing from the results of [ES85].

In the last section we present an open problem that comes out of this investigation. We exhibit a family $\{B_{2n+1} : n \in \mathbb{N}\}$ of graphs such that $|V(B_n)| = \Omega(2^n)$, whose vertices admit succint $O(n)$ descriptions. The results of Section 3 show that these graphs are bipartite, because they contain no odd cycle. However there is no known polynomial algorithm which will input a vertex of B_n and output a colour, and be guaranteed never to output the same colour for two adjacent vertices.

2 The Algorithm

Let G, H be given connected graphs, and $c : G \times H \to \{0,1,2\}$ a proper 3-colouring. We find a 3-colouring c' of G or of H with the following steps.

Step 1 Find an odd cycle $\{u_0, u_1, \ldots, u_{2m}\}$ in G; if none exists, output a 2-colouring c' of G and stop. Similarly, find an odd cycle $\{v_0,v_1,\ldots,v_{2n}\}$ in H; if none exists, output a 2-colouring c' of H and stop.

Step 2 Evaluate the cardinality of $\{i : c(u_i, v_0) \neq c(u_{i+2}, v_0)\}$ (with addition modulo $2m + 1$). If it is odd, G will be 3-coloured by the procedure given in Step 3. Otherwise, the cardinality of $\{j : c(u_0, v_j) \neq c(u_0, v_{j+2})\}$ (with addition modulo $2n + 1$) will be odd, and H will be 3-coloured by the procedure given in Step 3.

Step 3 Without loss of generality, assume that $|\{i : c(u_i, v_0) \neq c(u_{i+2}, v_0)\}|$ is odd. Then the subgraph G' of G induced by $\{x \in V(G) : c(x, v_0) \neq c(x, v_1)\}$ is bipartite, with bipartition $\{A, B\}$. A proper 3-colouring c' of G is then defined by

$$c'(x) = \begin{cases} c(x, v_0) & \text{if } x \in A \\ c(x, v_1) & \text{otherwise.} \end{cases}$$

This completes the algorithm when both factors are connected. If G or H is disconnected, the colouring c' is found by repeating the above procedure on every connected component. The algorithm is obviously polynomial if it can be run as stated. In Section 3, we will show that one of the sets $\{i : c(u_i, v_0) \neq c(u_{i+2}, v_0)\}$ and $\{j : c(u_0, v_j) \neq c(u_0, v_{j+2})\}$ of Step 2 has an odd cardinality, and that the graph G' of Step 3 is bipartite. Given these facts, it is not hard to show that the 3-colouring c' of Step 3 is indeed proper, using the fact that c is a proper 3-colouring of $G \times H$: Let $[x, y]$ be an edge of G. If both x and y are in G', then without loss of generality we can assume that $x \in A, y \in B$ so that $c'(x) = c(x, v_0) \neq c(y, v_1) = c'(y)$. Otherwise we can assume that x is not in G', so that $c'(x) = c(x, v_0) = c(x, v_1)$ is different from both $c(y, v_0)$ and $c(y, v_1)$.

3 Exponential Graphs

Following El-Zahar and Sauer, the "graph of 3-colourings of a n-cycle" is the graph $\mathcal{C}_3(C_n)$ whose vertices are all the functions $f : \mathbb{Z}_n \to \mathbb{Z}_3$, where two functions f, g are joined by an edge if $f(i) \neq g(i+1)$ and $g(i) \neq f(i+1)$ for all $i \in \mathbb{Z}_n$. The *parity* of a vertex f of $\mathcal{C}_3(C_n)$ is the parity of the cardinality of the set $\{i \in \mathbb{Z}_n : f(i-1) \neq f(i+1)\}$.

Given two graphs G, H, a proper 3-colouring c of $G \times H$ and an odd cycle $\{v_0, v_1, \ldots, v_{2n}\}$ in H, we can associate to each vertex u of G an element f_u of $\mathcal{C}_3(C_{2n+1})$, defined by $f_u(i) = c(u, v_i), i = 0, \ldots, 2n$. The fact that c is a proper colouring implies that whenever u and v are adjacent in G, f_u and f_v are adjacent in $\mathcal{C}_3(C_{2n+1})$ (that is, the map $\phi : G \to \mathcal{C}_3(C_{2n+1})$ defined by $\phi(u) = f_u$ is a homomorphism). However, the set $\{(u, v_i) : i = 0, \ldots, 2n\}$ is independent in $G \times H$, whence f_u could be any element of $\mathcal{C}_3(C_{2n+1})$, and any colouring of G derived from c must use the global structure of G, which is reflected in the parity of the maps f_u corresponding to its vertices. We use the following results of El-Zahar and Sauer:

Lemma 3.1 ([ES85], Lemma 3.1). *Let $f \in \mathcal{C}_3(C_n)$. Then the cardinality of the set*

$$\{i : f(i-1), f(i), f(i+1) \text{ are all different }\}$$

has the same parity as f itself.

Lemma 3.2 ([ES85], Lemma 3.3). *Let f and g be connected by an edge of $\mathcal{C}_3(C_n)$. Then f, g have the same parity.*

Lemma 3.3 ([ES85], Lemma 3.4). *Let $V(C_{2m+1}) = \{u_0, u_1, \ldots, u_{2m}\}$ and $V(C_{2n+1}) = \{v_0, v_1, \ldots, v_{2n}\}$. Then for any proper 3-colouring c of $C_{2m+1} \times C_{2n+1}$, the parity of the induced map $f_{u_0} \in \mathcal{C}_3(C_{2n+1})$ is different from the parity of the induced map $f_{v_0} \in \mathcal{C}_3(C_{2m+1})$.*

Lemma 3.3 asserts that Step 2 of the algorithm of the previous section will indeed select a graph to colour. Lemma 3.2 shows that parity is an invariant of connected components of $\mathcal{C}_3(C_n)$. Let B_{2n+1} be the subgraph of $\mathcal{C}_3(C_{2n+1})$ induced by the set

$$V(B_{2n+1}) = \{f \in V(\mathcal{C}_3(C_{2n+1})) : f \text{ has even parity and } f(0) \neq f(1)\}.$$

Lemma 3.4. *For every $n \geq 1$, B_{2n+1} is bipartite.*

Proof. El-Zahar and Sauer provide the following argument in the proof of some other result. Suppose that $\{f_0, f_1, \ldots, f_{2m}\}$ is an odd cycle of B_{2n+1}. Consider the proper colouring c of $C_{2m+1} \times C_{2n+1}$ defined by $c(u_i, v_j) = f_i(j)$. Since $f_{v_0} = f_0$, f_{u_0} must be odd by Lemma 3.2. The map $f_{u_0} \in \mathcal{C}_3(C_{2m+1})$ is defined by $f_{u_0}(i) = f_i(0)$, and by Lemma 3.1 there exists an index i such that $f_{i-1}(0), f_i(0)$ and $f_{i+1}(0)$ are all distinct. We must then have $f_i(1)$ distinct from both $f_{i-1}(0)$ and $f_{i+1}(0)$, since f_i is adjacent to f_{i-1} and f_{i+1}. Thus $f_i(1) = f_i(0)$, which contradicts the fact that f_i is in B_{2n+1}. Therefore B_{2n+1} does not contain an odd cycle, and is bipartite. □

The graph G' defined in Step 3 of the algorithm of Section 2 is a subgraph of B_{2n+1}, hence it is bipartite by the previous lemma. This completes our verification of the correctness of the algorithm of Section 2.

4 Explicit Colourings of B_{2n+1}

We conclude this note with a problem that we have not been able to solve.

Problem 4.1. Does there exist a sequence $(c_{2n+1} : B_{2n+1} \to \mathbb{Z}_2)_{n\geq 1}$ of proper colourings and an algorithm which is polynomial in n, and which inputs a vertex $f \in V(B_{2n+1})$ and outputs $c_{2n+1}(f)$?

It can be shown that B_{2n+1} has at least $4 \cdot 3^{2n-2}$ vertices, though its vertices admit succint representations as vectors in $\mathbb{Z}_3^{2n+1}$, and adjacency between two vertices can be decided in time $O(n)$. The proof of Lemma 3.4 is nonconstructive and only shows that B_{2n+1} cannot contain an odd cycle. Hence a family of colourings $(c_{2n+1} : B_{2n+1} \to \mathbb{Z}_2)_{n\geq 1}$ can be defined by deciding that the lexicographically minimal element of each connected component of B_{2n+1} gets colour 0. We then find the colour of an arbitrary $f \in V(B_{2n+1})$ as follows. We first find the lexicographically minimal $f' \in V(B_{2n+1})$ which is reachable from f. We then decide if f' also reachable from f in the graph D_{2n+1}, where $V(D_{2n+1}) = V(B_{2n+1})$ and $[g, h] \in E(D_{2n+1})$ if g and h have a common neighbour in B_{2n+1}. If f' is reachable from f in D_{2n+1}, then f gets colour 0, and otherwise f gets colour 1. Reachability in B_{2n+1} and D_{2n+1} is now known to be decidable in log-space in terms of $|V(B_{2n+1})|$, that is, polynomial space in terms of n. However the computation of $c_{2n+1}(f)$ described above still takes exponential time.

The situation of the cubes closely parallels that of the graphs B_{2n+1} discussed here: The cube Q_n is the graph whose vertices are the 2^n vectors in $\mathbb{Z}_2^n$, where two vectors are adjacent when they differ in exactly one coordinate. It is well known that the cubes are bipartite, and in fact, there is a family of explicit colourings $(c_n : Q_2 \to \mathbb{Z}_2)_{n\geq 1}$ defined by $c_n(x_1, \ldots, x_n) = x_1 + \cdots + x_n$. Perhaps there exist similar easily computable explicit colourings of the graphs B_{2n+1}, but none have been found so far.

References

[ES85] M. El-Zahar, N. Sauer, The chromatic number of the product of two 4-chromatic graphs is 4, *Combinatorica* **5** (1985), 121–126.

[Hed66] S. H. Hedetniemi, Homomorphisms of graphs and automata, University of Michigan Technical Report 03105-44-T, 1966.

Recent Developments in Circular Colouring of Graphs

Xuding Zhu*

Department of Applied Mathematics, National Sun Yat-sen University, Kaohsiung, 80424 Taiwan
zhu@math.nsysu.edu.tw

Summary. The study of circular chromatic number $\chi_c(G)$ of a graph G, which is a refinement of its chromatic number, has been very active in the past decade. Many nice results are obtained, new techniques are developed, and connections to other fields are established. This paper presents a glimpse of the recent progress on this subject. Besides presenting the results, some of the ideas and tools in the proofs are explained, although no detailed proofs are contained.

Keywords. Circular chromatic number, circular chromatic index, circular perfect graphs, circular flow number, graph homomorphism.

1 Definitions

The circular chromatic number of a graph can be defined in a few different but equivalent ways. We first list some of the commonly used definitions.

Circular r-colouring. Suppose $G = (V, E)$ is a graph and $r \geq 1$ is real number. A *circular r-colouring* of G is a mapping $f : V \to [0, r)$ such that for any edge xy of G, $1 \leq |f(x) - f(y)| \leq r - 1$. We say a graph G is *circular r-colourable* if G has a circular r-colouring. The *circular chromatic number* $\chi_c(G)$ of G is defined as

$$\chi_c(G) = \inf\{r : G \text{ is circular } r\text{-colourable.}\}$$

It is known [BZ98, BH90, Vin88, Zhu01c] that the infimum in the definition above is always attained (even if G is an infinite graph), and hence can be replaced by the minimum. If $r = k$ is an integer, then a k-colouring is a circular k-colouring. Conversely, if f is a circular k-colouring of G then $g(v) = \lfloor f(v) \rfloor$ defines a k-colouring of G. So a graph G is circular k-colourable if and only if

* This research was partially supported by the National Science Council under grant NSC93-2115-M-110-004.

G is k-colourable. For this reason, a circular r-colouring of a graph G is usually simply called an r-colouring of G. Instead of saying G is circular r-colourable, we usually simply say G is r-colourable. It is obvious that if $r' \geq r$ and G is r-colourable then G is r'-colourable. This implies that for any graph G,

$$\chi(G) - 1 < \chi_c(G) \leq \chi(G). \tag{1}$$

Inequality (1) shows that $\chi_c(G)$ contains more information about the structure of the graph G than $\chi(G)$ does. We say $\chi_c(G)$ is a refinement of $\chi(G)$ and $\chi(G)$ is an approximation of $\chi_c(G)$. If G is a finite graph, then $\chi_c(G) = p/q$ for some integers p, q (not necessarily relatively prime), such that G has a cycle of length p (unless G is a forest) and every vertex of G is contained in an independent set of size at least q. So if G has n vertices, then $\chi_c(G) = p/q$ for some $p \leq n$. Hence (1) can be improved to

$$(\chi(G) - 1)(1 + \frac{1}{n-1}) \leq \chi_c(G) \leq \chi(G). \tag{1'}$$

In a circular r-colouring of a graph G, the "colour set" consists of all the real numbers in the interval $[0, r)$. A better way of picturing the colour set is to identify 0 and r of the interval $[0, r]$ into a single point to obtain a circle of perimeter r. We denote this circle by S^r. So the colour set is the set of points on the circle S^r. For two points $a, b \in S^r$, the distance between a, b, denoted by $|a - b|_r$, is the length of the shorter arc of S^r connecting a and b. For a real number x and a positive real number r, we denote by $[x]_r$ the remainder of x dividing r, i.e., $[x]_r \in [0, r)$ is the unique number for which $x - [x]_r$ is a multiple of r. Then

$$|a - b|_r = \min\{[a - b]_r, [b - a]_r\} = \min\{|a - b|, r - |a - b|\}.$$

By this notation, a circular r-colouring of G is a mapping f which assigns to each vertex x of G a point $f(x) \in S^r$ such that for any edge xy of G, $|f(x) - f(y)|_r \geq 1$. This interpretation explains the name "circular colouring" and "circular chromatic number".

(p, q)-colouring. Another useful definition of circular chromatic number uses the concept of (p, q)-colouring, where only finitely many colours are used. Suppose $p \geq q$ are positive integers. A (p, q)-colouring of a graph $G = (V, E)$ is a mapping $f : V \to \{0, 1, \ldots, p - 1\}$ such that for any edge xy of G, $q \leq |f(x) - f(y)| \leq p - q$. If f is a (p, q)-colouring of G, then the mapping $g(x) = f(x)/q$ defines a $\frac{p}{q}$-colouring of G. Conversely, if g is a $\frac{p}{q}$-colouring of G, then the mapping $f(x) = \lfloor g(x)q \rfloor$ defines a (p, q)-colouring. Therefore, for any graph G,

$$\chi_c(G) = \inf\{\frac{p}{q} : G \text{ has a } (p, q)\text{-colouring}\}.$$

If G is finite, then the infimum can be replaced by minimum. But infinite graphs G can have $\chi_c(G)$ equal to irrationals and the infimum cannot be replaced by the minimum.

Graph homomorphism. Suppose G, H are graphs. A *homomorphism* of G to H is a mapping $f : V(G) \to V(H)$ such that for every edge xy of G, $f(x)f(y)$ is an edge of H. Two graphs G and H are *homorphically equivalent*, written as $G \sim H$, if each admits a homomorphism to the other. For positive integers $p \geq 2q$, let $K_{p/q}$ be the graph with vertex set $\{0, 1, \ldots, p-1\}$ in which ij is an edge if and only if $q \leq |i-j| \leq p-q$. Then a (p,q)-colouring of a graph G is simply a homomorphism of G to $K_{p/q}$.

Graph homomorphisms provide a unified language and useful tool for the study of different graph colouring problems [HT97, HN04]. We write $G \preceq H$ if there is a homomorphism from G to H. Let $\mathcal{G}$ denote the set of (isomorphic classes of) graphs. Then $(\mathcal{G}/\sim, \preceq)$ is a partial order. The set $\mathcal{Z}_{\mathcal{G}} = \{K_1, K_2, \ldots\}$ of complete graphs form an infinite increasing chain in this partial order. This infinite increasing chain provides a scale that measures the 'colourability' of graphs. For a graph G, the chromatic number of G is the least k for which $G \preceq K_k$. The graphs $K_{p/q}$ play a special role in the study of circular colourings, and are called *circular complete graphs*. Let $\mathcal{Q}_{\mathcal{G}} = \{K_{p/q} : p \geq 2q, (p,q) = 1\} \cup \{K_1\}$. Then $\mathcal{Q}_{\mathcal{G}}$ is a superset of $\mathcal{Z}_{\mathcal{G}}$, and provides a finer scale that measures the colourability of graphs [Zhu99c]. The circular chromatic number of G can be defined as

$$\chi_c(G) = \inf\{p/q : G \preceq K_{p/q}\}.$$

Orientation. Another point of view of circular chromatic number is reflected in a result of Hoffman [Hof60], relating the chromatic number of a graph G to orientations of G. An orientation D of a graph G is obtained from G by assigning to each edge an orientation. An edge with a given orientation is called an *arc*, and the set of arcs of D is denoted by $A(D)$. Suppose D is an orientation of G and C is a cycle of G. The *imbalance* of a cycle C of G with respect to D is $\mathrm{Imb}_D(C) = \max\{|C|/|C^+|, |C|/|C^-|\}$, where C^+ and C^- are the sets of forward arcs and backward arcs of C, respectively. The *Cycle Imbalance of* D is defined as $\mathrm{CycImb}(D) = \sup\{\mathrm{Imb}_D(C) : C \text{ is a cycle of } G\}$. Hoffman's Lemma says that for any graph G which is not a forest ,

$$\chi(G) = \inf\{\lceil \mathrm{CycImb}(D) \rceil, D \text{ is an acyclic orientations of } G\}.$$

It is proved in [GTZ98] (cf. [BH02]) that if the ceiling function is omitted, then one obtains the circular chromatic number of G, i.e.,

$$\chi_c(G) = \inf\{\mathrm{CycImb}(D), D \text{ is an acyclic orientations of } G\}.$$

So to prove a graph G has $\chi_c(G) \leq p/q$, it suffices to have an orientation D in which each cycle C has $|C|/|C^+| \leq p/q$ and $|C|/|C^-| \leq p/q$. Indeed, it is shown in [Zhu02b] that it suffices to check those cycles C for which $[q|C|]_p \in \{1, 2, \ldots, 2q-1\}$.

Tension. Suppose G is a graph and D is an orientation of G. A *tension* is a mapping $f : A(D) \to \mathbb{R}$ which assigns to each arc e of D a real number $f(e)$ such that for each cycle C of G,

$$\sum_{e \in C^+} f(e) = \sum_{e \in C^-} f(e).$$

For a real number $r \geq 2$, an *r-tension* of G is a tension f such that for each arc e, $1 \leq |f(e)| \leq r - 1$. Each r-colouring ϕ of G corresponds to an r-tension f of G defined as $f(e) = \phi(y) - \phi(x)$, where $e = (x, y)$ is an arc. Conversely, if we have an r-tension f of G, then let x^* be a fixed vertex of G, and for each vertex x of G let W_x be an arbitrary x^*-x-walk. Then $\phi(x) = [\sum_{e \in W_x^+} f(e) - \sum_{e \in W_x^-} f(e)]_r$ defines an r-colouring of G. Therefore,

$$\chi_c(G) = \min\{r : \text{ there is an } r\text{-tension of} G\}.$$

The different points of view of looking at the circular chromatic number of graphs show that the parameter $\chi_c(G)$ is a very natural refinement of $\chi(G)$.

2 Circular Colouring and Periodic Scheduling

Graph colouring is an ideal model for various scheduling problems. If the scheduling is periodic, then it is very likely that circular colouring of graphs provides a more accurate model. In [Zhu01c], a traffic light problem (assigning the green light phases to the traffic flows) is used to motivate the definition of circular chromatic number. The traffic light problem is a typical periodic scheduling problem. However, such problems are usually not very complicated, and are usually solved by experience, without explicitly using the graph model. On the other hand, there are many periodic scheduling problems that are very complicated, and mathematical models are needed to analyze them. Computer science is a rich source for periodic scheduling problems, and sometimes circular colouring of graphs can be used in finding optimal solutions in such problems.

One problem studied extensively by computer scientists is the concurrency of *heavily loaded resource sharing systems*. Let V be a set of processes, and D be a set of data files. Each process x has access to a set $D(x) \subseteq D$ of data files. When a process x operates, it accesses all the files in $D(x)$. Therefore if processes x and y share a common data file, then x and y cannot operate at the same time. To ensure fairness, if x and y share a common data file, then x and y must alternate their turns to operate. In a heavily loaded resource sharing system, all the processes are constantly demanding access to all resources that they use. Subject to the constraints of fairness and that processes sharing a resource cannot operate at the same time, our task is to schedule the operating time of processes efficiently.

This problem is modeled by a graph, where each vertex represents a process, and two vertices are adjacent if the corresponding processes share a resource. A *scheduling* of G is a mapping f which assigns to each vertex x of G a subset $f(x)$ of $\{0, 1, \ldots\}$. The interpretation is that if $i \in f(x)$ then x operates at time i. A scheduling f of G is *valid* if for any edge xy of G,

$f(x) \cap f(y) = \emptyset$. Moreover, if $i < i'$ and $i, i' \in f(x)$, then there is an integer $j \in f(y)$ such that $i < j < i'$. The efficiency $\sigma(f)$ of the scheduling f is $\sigma(f) = \liminf_{n\to\infty} \sum_{i=0}^{n} \frac{|f^{-1}(i)|}{n|V(G)|}$, which is the average portion of processes in operation. The *concurrency* of a graph G, denoted by $\xi^*(G)$, is defined as

$$\xi^*(G) = \sup\{\sigma(f) : f \text{ is a valid scheduling of } G\}.$$

For example, if $G = C_5$ is the 5-cycle with vertex set $\{v_1, v_2, v_3, v_4, v_5\}$ and edge set $\{v_1v_2, v_2v_3, v_3v_4, v_4v_5, v_5v_1\}$, then the mapping $f(v_1) = f(v_3) = \{3i : i = 0, 1, \ldots\}$, $f(v_2) = f(v_4) = \{3i+1 : i = 0, 1, \ldots\}$ and $f(v_5) = \{3i+2 : i = 0, 1, \ldots\}$ is a valid scheduling with efficiency $1/3$. In general, if G is a k-chromatic graph and c is a k-colouring of G, then $f(x) = \{i : [i]_k = c(x)\}$ is a valid scheduling with efficiency $1/k$.

One method for finding an optimal scheduling is developed by computer scientists. The method is called the *edge reversal method.* Given an acyclic orientation of G, the scheduling induced by the orientation is obtained by repeatedly applying the following step: Let all the sinks operate and reverse the orientation of those edges incident to sinks.

Consider the 5-cycle C_5, with initial orientation as in Figure 1(a). Then

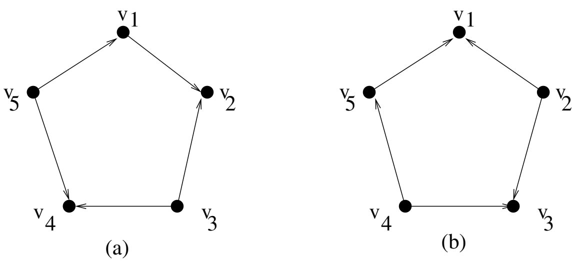


Fig. 1. The initial and the second orientation of C_5

v_2, v_4 operate on the first round and after the operation, the orientation will be changed to the one in Figure 1(b). Repeating the process, v_1, v_3 operate in the second round, v_2, v_5 the third round, v_1, v_3 the 4th round, v_2, v_5 the 5th round, and so on. The efficiency of this scheduling is $2/5$, better than the scheduling derived from the 3-colouring of C_5.

It is proved by Barbosa and Gafni [BG89] that given an acyclic orientation D of G, the efficiency of the scheduling derived from D is equal to the reciprocal of the imbalance of D. Moreover, for any graph G, there is an optimal scheduling of G which is derived from an acyclic orientation of G. In other words, $1/\xi^*(G) = \max\{\mathrm{Imb}(D) : D \text{ is an acyclic orientation of } G\}$. This implies that for any graph G,

$$\chi_c(G) = 1/\xi^*(G).$$

The study of circular chromatic number of graphs by graph theorists is apparently originated from the paper [Vin88] by Vince, published in 1988. It follows from the result above that $\xi^*(G)$ is just the reciprocal of $\chi_c(G)$. The parameter $\xi^*(G)$ is defined in [Bar86], and the results were consequently published in conference and journal [BG89]. The edge reversal method has been studied by computer scientists in many papers, however, the parameter $\xi^*(G)$ did not attract as much attention as $\chi_c(G)$.

A *fractional colouring* of G is a mapping f which assigns to each independent set I of G a non-negative weight $f(I)$ so that for each vertex x, $\sum_{x \in I} f(I) = 1$. If $\sum f(I) = r$ (where the summation is over all independent sets I of G), then f is called an *r-fractional colouring* of G. The *fractional chromatic number* of G is defined as $\chi_f(G) = \inf\{r : G$ has an r-fractional colouring $\}$. It is known to both computer scientists and graph theorists [Bar02, Vin88] that for any graph G, $\chi_f(G) \leq 1/\xi^*(G) \leq \chi(G)$. However, the relation that $\lceil \chi_c(G) \rceil = \chi(G)$ is not obvious from the definition of $\xi^*(G)$, and the question whether $\xi^*(G)$ determines $\chi(G)$ was posed as an open problem in [BG87].

Although computer scientists and graph theorists are both interested in $\chi_c(G)$, they studied the parameter independently, using different languages, and without knowing the existence of the other side. It is until recently that the connection between the two sides has been revealed [YZ05]. In [YZ05], a formula is given that directly transform a periodic valid scheduling of a graph G into a circular colouring of G and vice versa.

Some other problems studied in computer sciences as well as in operations research can also be modeled as circular colouring of graphs, or circular colouring of edge weighted digraphs.

3 Circular Colouring of Digraphs

Let S^r be a circle of perimeter r as defined before. For two points p, p' of S^r, let $d(p, p') = [p' - p]_r$ be the length of the arc of S^r from p to p' along the clockwise direction.

An r-colouring of a digraph G is a mapping $f : V(G) \to S^r$ such that for each arc (x, y) of G, $d(f(x), f(y)) \geq 1$. We say G is r-colourable if there is an r-colouring of G. The circular chromatic number of a digraph G is defined as

$$\chi_c(G) = \inf\{r : G \text{ is } r\text{-colourable}\}.$$

By viewing an undirected graph G as a symmetric digraph, in which each edge $e = xy$ of G corresponds to two opposite arcs (x, y) and (y, x), the above definition of an r-colouring of the symmetric digraph is equivalent to the original definition of an r-colouring of the undirected graph. So the r-colouring of digraphs, introduced in [BFJKM04], is a very natural generalization of the r-colouring of undirected graphs. Many results concerning r-colouring of undirected graphs generalize to r-colouring of digraphs without difficulties. There are also differences between undirected graphs and digraphs. One difference is that the infimum in the definition of the circular chromatic number of a digraph cannot be replaced by the minimum. For example, if G is an acyclic digraph, then for any $r > 1$, there is an r-colouring of G. So $\chi_c(G) = 1$. However, there is no 1-colouring of G. One may argue that this reflects one subtle aspect of the definition of distance between points on the circle S^r. Given a point p on S^r, what should be the distance $d(p,p)$? Should it be 0? Or should it be r? It is not a good idea to have $d(p,p) = r$, for otherwise one may colour all the vertices of a digraph G by the same colour to conclude that all digraphs are 1-colourable. On the other hand, we have $\lim_{p' \to p} d(p,p') = r$, if the limit is taken in the appropriate direction. To take this into consideration, the following definition is introduced in [BFJKM04]:

A weak circular r-colouring of a digraph G is a mapping $f : V(G) \to S^r$ such that for each arc (x,y), either $f(x) = f(y)$ or $d(f(x), f(y)) \geq 1$. Moreover, for any point p of S^r, $f^{-1}(p)$ induces an acyclic subgraph of G.

Observe that the definition of weak circular colouring applied to symmetric digraphs is also equivalent to the original definition. Using the weak circular r-colourability of digraphs, the circular chromatic number of a digraph G can be proved to be

$$\chi_c(G) = \min\{r : \ G \text{ is weak circular } r\text{-colourable}\}.$$

The circular chromatic number of an undirected graph is either 1 or at least 2. But for any rational $r \geq 1$, there is a finite digraph G with $\chi_c(G) = r$.

Theorem 3.1. *Suppose $p \geq q$. Let $\vec{K}_{p/q}$ be the digraph with vertex set $\{0, 1, \ldots, p-1\}$ in which (i,j) is an arc if and only if $[j-i]_p \geq q$. Then $\chi_c(\vec{K}_{p/q}) = p/q$. In particular, for every rational $p/q \geq 1$, there exists a finite digraph with circular chromatic number p/q.*

Indeed, the digraphs $\vec{K}_{p/q}$ play the same role in the study of circular colouring of digraphs as the circular complete graphs $K_{p/q}$ in the study of circular colouring of undirected graphs. We define an *acyclic homomorphism* of a digraph G to a digraph G' as a mapping $f : V(G) \to V(G')$ such that for each arc (x,y) of G, either $f(x) = f(y)$ or $(f(x), f(y))$ is an arc of G'. Moreover, for each vertex v of G', $f^{-1}(x)$ induces an acyclic sub-digraph of G. Then we have the following result.

Theorem 3.2. *A digraph G has circular chromatic number at most p/q if and only if there exists an acyclic homomorphism of G to $\vec{K}_{p/q}$.*

Corollary 3.3. *If $p/q \leq p'/q'$ and G admits an acyclic homomorphism to $\vec{K}_{p/q}$ then G admits an acyclic homomorphism to $\vec{K}_{p'/q'}$.*

It is proved in [Vin88] that if G is a finite undirected graph on n vertices then $\chi_c(G) = p/q$ for some $p \leq n$. The same conclusion holds for the circular chromatic number of digraphs. Given a weak circular r-colouring c of a digraph G, a cycle $C = (v_1, v_2, \ldots, v_k, v_1)$ in the underlying graph of G is called a *tight cycle* if for each i, if (v_i, v_{i+1}) is an arc of C then $d(c(v_i), c(v_{i+1})) = 1$, otherwise (v_{i+1}, v_i) is an arc of C and $c(v_i) = c(v_{i+1})$, where additions in the indices are modulo k. The following result is a generalization of a result in [Gui93], and is a strengthening of a result in [Moh03].

Theorem 3.4. *A digraph G has $\chi_c(G) = r$ if and only if there is a weak circular r-colouring of G, and moreover, every weak circular r-colouring of G has a tight cycle.*

If C is a tight cycle of a weak circular r-colouring of G, then the weight $a(C)$ of C is the number of *forward edges* of C, i.e., number of indices i for which (v_i, v_{i+1}) is an arc. It follows from the definition of tight cycle that the weight $a(C)$ is a multiple of r, say $a(C) = qr$ for some positive integer q. As $a(C) = p$ is an integer less than or equal to $|C|$, we conclude that for any digraph G, $\chi_c(G) = p/q$ for some $p \leq |C| \leq |V(G)|$.

The definition of circular chromatic number of digraphs leads to the definition of *chromatic number* $\chi(G)$ of a digraph G to be the minimum integer k such that $V(G)$ can be partitioned into k acyclic subsets. In other words, $\chi(G)$ is the minimum integer r for which G has a weak circular r-colouring. Therefore for any digraph G we have

$$\chi(G) - 1 < \chi_c(G) \leq \chi(G).$$

In case the digraph G is symmetric, then $\chi(G)$ coincides with the definition of chromatic number of undirected graphs.

Hell and Nešetřil [HN89] proved that if H is non-bipartite, then it is NP-complete to decide if an arbitrary graph G admits a homomorphism to H. As a consequence, for any $r > 2$, it is NP-complete to determine if $\chi_c(G) \leq r$. For digraphs, it is easy to see that if H is acyclic, then it is polynomial to decide if an arbitrary digraph G admits an acyclic homomorphism to H. Feder, Hell and Mohar [FHM03] proved that if H is not acyclic, then it is NP-complete to decide if an arbitrary digraph admits an acyclic homomorphism to H. It follows from this result that for any $r > 1$, it is NP-complete to decide if an arbitrary digraph G satisfies $\chi_c(G) \leq r$.

Given a digraph G, let $\alpha(G)$ be the maximum size of a subset of $V(G)$ that induces an acyclic sub-digraph. Then we have $\chi(G) \geq |V(G)|/\alpha(G)$ for any digraph G. By a probabilistic argument, it is proved in [BFJKM04] that for any integer ℓ, there is a digraph G on n vertices with a digraph with digirth (i.e., the length of a shortest directed cycle) at least ℓ, and with $\alpha(G) \leq O(n^{1-\theta} \ln n)$ for a positive $\theta < 1/\ell$. So if n is large enough, then $\chi(G) \geq k$ for any given constant k. This implies the following result [BFJKM04]:

Theorem 3.5. *For any integers k, ℓ, there is a digraph G with digirth at least ℓ and with circular chromatic number at least k.*

Almost all problems concerning the circular chromatic number of undirected graphs can be asked in terms of digraphs. The questions of possible values of the circular chromatic number of undirected planar graphs is answered by Moser [Mos97] and this author [Zhu99e, Zhu99b].

Theorem 3.6. *There is a finite undirected planar graph G with $\chi_c(G) = r$ if and only if $r = 1$ or r is a rational and $2 \le r \le 4$.*

Since, unlike the case for undirected graphs, the circular chromatic number of a digraph can be strictly between 1 and 2, one may wonder if every rational number between 1 and 2 is the circular chromatic number of a planar digraph. Soh [Soh05] showed that the answer is yes.

Theorem 3.7. *There is a finite planar digraph G with $\chi_c(G) = r$ if and only if r is a rational and $1 \le r \le 4$.*

The maximum chromatic number of planar graphs (the Four Colour Problem) plays an important role in graph colouring theory. It is natural to ask what is the maximum acyclic chromatic number of an orientation of a planar graph. The following conjecture is proposed in [BFJKM04]:

Conjecture 3.8. *If G is a planar digraph without 2-cycles, then the acyclic chromatic number of G is at most 2. I.e., $V(G)$ can be partitioned into $V_1 \cup V_2$ such that each V_i induces an acyclic sub-digraph of G.*

A more general version of circular colouring of digraphs is to consider edge weighted digraphs. Let $G = (V, E)$ be a digraph and $c : E \to \mathbb{R}^+ \cup \{0\}$ be the edge weights. An r-colouring of G is a mapping $f : V \to [0, r)$ such that for any arc $(x, y) \in E$, $[f(y) - f(x)]_r \ge c_{xy}$ (we use c_{xy} to denote the weight of arc (x, y)). The r-colouring of digraphs is the special case that $c_{xy} = 1$ for all $(x, y) \in E$. Basic properties of circular colouring of edge weighted digraphs is studied in [Moh03] and it is shown there that circular colouring of edge weighted digraphs generalizes the earlier concept of circular colouring of vertex weighted graphs [DZ96].

Circular colouring of edge weighted digraphs provides a model for parallel computations. Let $G = (V, E)$ be a digraph, and $c : E \to \mathbb{R}^+$ be the edge weight function. Let $T : E \to \mathbb{N}$ be a mapping which assigns to each arc (u, v) a number T_{uv} of *tokens*. The triple (G, c, T) is called a *timed marked graph.* A timed marked graph (G, c, T) can be used to model the data movement in parallel computations. A vertex represents a task, an arc (u, v) represents a data channel. A token on arc (u, v) represents an input from u to v, and T is the initial assignment of tokens. When a vertex operates, it consumes one token from each of its in-arcs, and produces a token for each of its out-arcs. The weight c_{uv} represents the time required by task vertex u to place the result of its operation on (u, v). So if u operates at time t, then at time $t + c_{uv}$, a token is placed on (u, v) and becomes available to v. A scheduling for the timed marked graph determines, for each vertex v, the time pulses at which

v operates. The scheduling is *admissible* if whenever a vertex v operates, each in-arc of v has at least one token available. Computer scientists are interested in periodic schedulings, in which each vertex v is assigned a single time pulse $\phi(u)$, and it operates at time pulses $\phi(u) + pk$ for $k = 0, 1, \ldots$ Here p is the period. An initial marking T of an edge weighted digraph (G, c) is *good* if for each directed cycle C, $\sum_{(u,v) \in C} T_{uv} > 0$ and for each arc (u, v), $T_{uv} + T_{vu} = 1$. Connection of circular colouring and periodic scheduling of timed marked graphs is studied in [Yeh05], where the following result is proved.

Theorem 3.9. *An edge-weighted symmetric digraph (G, c) has a circular p-colouring if and only if there is a good initial marking T for (G, c) for which the timed marked graph (G, c, T) admits a periodic admissible schedule with period p.*

4 Circular Chromatic Index

Given a graph G, the *line graph* of G, denoted by $L(G)$, has vertex set $E(G)$, in which e, e' are adjacent if e and e' have a common end-vertex. The *chromatic index* $\chi'(G)$ of G is defined as $\chi'(G) = \chi(L(G))$ and the *circular chromatic index* $\chi'_c(G)$ of G is defined as $\chi'_c(G) = \chi_c(L(G))$. So we have

$$\chi'(G) - 1 < \chi'_c(G) \leq \chi'(G).$$

If G is connected and $\Delta(G) = 2$, then G is either a cycle or a path. This implies that either $\chi'_c(G) = 2$ or $\chi'_c(G) = 2 + \frac{1}{k}$ for some positive integer k. Since graphs G with $\Delta(G) \geq 3$ have $\chi'_c(G) \geq 3$, 'most' of the rational numbers in the interval $(2, 3)$ are not the circular chromatic index of any graph.

Suppose G is a connected graph with $\Delta(G) = 3$. Then $3 \leq \chi'_c(G) \leq 4$. A natural question is that what are the possible values of the circular chromatic index of such a graph. It is well-known that the Four Color Theorem is equivalent to the statement that every 2-edge connected cubic planar graph G has $\chi'_c(G) = 3$. For nonplanar 2-edge connected cubic graphs, Jaeger [Jae88] proposed the following conjecture (the Petersen Coloring Conjecture):

Conjecture 4.1. *If G is a 2-edge connected cubic graph, then one can colour the edges of G, using the edges of the Petersen graph as colours, in such a way that any three mutually adjacent edges of G are coloured by three edges that are mutually adjacent in the Petersen graph.*

Since the Petersen graph has circular chromatic index $11/3$, Conjecture 4.1 would imply that every 2-edge connected cubic graph G has $\chi'_c(G) \leq 11/3$. This consequence of Conjecture 4.1 is confirmed in [AGGHTZ05].

Theorem 4.2. *The circular chromatic index of every 2-edge connected cubic graph G (parallel edges allowed) is less than or equal to $11/3$.*

Indeed, a more general result is proved in [AGGHTZ05].

Theorem 4.3. *Suppose G is 2-edge connected and has maximum degree 3 (parallel edges are allowed). If $G \neq H_1, H_2$ (where H_1, H_2 are the graphs in Figure 2), then $\chi'_c(G) \leq 11/3$.*

Fig. 2. (a) The graph H_1 (b) The graph H_2

Corollary 4.4. *If G is graph (parallel edges allowed) of maximum degree 3 and G does not contain H_1 or H_2 as a subgraph, then $\chi'_c(G) \leq 11/3$.*

It is easy to verify that $\chi'_c(H_1) = \chi'_c(H_2) = 4$. Since graphs G with $\Delta(G) \geq 4$ have $\chi'_c(G) \geq 4$. Therefore we have the following:

Corollary 4.5. *There is no graph G with $11/3 < \chi'_c(G) < 4$.*

So the interval $(11/3, 4)$ is a gap for the circular chromatic indexes of graphs, and we do not know if there are other gaps. I propose the following conjecture.

Conjecture 4.6. *Let Ω be the set of all the circular chromatic indexes of graphs, i.e., $\Omega = \{\chi'_c(G) : G$ is a finite graph$\}$. There is no bounded strictly increasing infinite sequence in Ω.*

This conjecture is strong, and implies that there are gaps everywhere for the circular chromatic indexes of graphs. Here is a weaker conjecture.

Conjecture 4.7. *For each integer n, there is an $\epsilon_n > 0$ such that there is no graph G with $\chi'_c(G) \in (n - \varepsilon_n, n)$.*

If $n = 2, 3, 4$, then ε_n exist and can be chosen as $\varepsilon_n = 1/(n-1)$. It might be true that ε_n can be chosen as $1/(n-1)$ for $n \geq 5$ as well, as conjectured in [AGGHTZ05, KKŠ05] (see also `http://www.math.nsysu.edu.tw/~zhu/open-problems`).

It is conjectured by Jaeger and Swart [JW80] that every 2-edge connected cubic graph of large girth admits a nowhere zero 4-flow. This is equivalent to say that 2-edge connected cubic graph of large girth are 3-edge colourable.

This conjecture (called the Girth Conjecture) is refuted by Kochol [Koc96]. However, if one considers circular chromatic index instead of chromatic index, then the Girth Conjecture is almost true. The circular chromatic index of cubic graphs of large girth is studied by Kaiser, Král and Škrekovski [KKŠ05]. They proved the following result.

Theorem 4.8. *For each $\varepsilon > 0$ there is an integer $n = n(\varepsilon)$ such that any cubic graph of girth at least n has $\chi'_c(G) < 3 + \varepsilon$.*

This result is generalized in [KKŠZ05].

Theorem 4.9. *For each $\varepsilon > 0$ and any integer Δ, there is an integer $n = n(\Delta, \varepsilon)$ such that if G is a graph with maximum degree Δ and with girth at least n has $\chi'_c(G) < \Delta + \varepsilon$.*

The circular chromatic index of some special classes of graphs have been studied in a few papers [GKNT06, HK04, Mos95, Nad06, Jhan05, WZ06]. Observe that if G has no parallel edges, then by Vizing Theorem, $\chi'(G) = \Delta(G)$ or $\Delta(G) + 1$. Graphs G with $\chi'(G) = \Delta(G)$ are called *class 1* graphs, and graphs with $\chi'(G) = \Delta(G) + 1$ are called *class 2* graphs. If G is class 1, then since $\omega(L(G)) = \Delta(G)$, we conclude that $\chi'_c(G) = \Delta(G)$. Thus for the study of the circular chromatic index of graphs, we are interested in class 2 graphs.

Cyclically 4-edge connected cubic graphs of class 2 are called *snarks*. One well-known infinite family of snarks is the family of flower snarks. The *flower snark* J_{2k+1} is obtained from the disjoint union of two cycles, $(a_0, a_1, \dots, a_{2k})$ and $(c_0, d_1, c_2, d_3, \dots, c_{2k}, d_0, c_1, d_2, c_3, \dots, c_{2k-1}, d_{2k})$, by adding vertices b_i and edges a_ib_i, c_ib_i, d_ib_i for $i = 0, 1, 2, \dots, 2k$. The circular chromatic index of flower snarks are completely determined [GKNT06].

Theorem 4.10. *The circular chromatic index of flower snarks are as follows: $\chi'_c(J_3) = 7/2$, $\chi'_c(J_5) = 17/5$ and for $k \geq 3$, $\chi'_c(J_{2k+1}) = 10/3$.*

The circular chromatic index of the Cartesian product of graphs is studied in [WZ06]. In particular, the circular chromatic index of the Cartesian product of two odd cycles is estimated.

Theorem 4.11. *For any $m \geq k$,*

$$\chi_c(C_{2k+1} \square C_{2m+1}) \geq 4 + \frac{1}{\lfloor (3(2k+1))/4 \rfloor}.$$

Moreover, if $m \geq 3k + 1$, then equality holds.

5 Graph Products

Suppose $G = (V, E)$ and $G' = (V', E')$. The *categorical product* $G \times G'$ of G and G' has vertex set $V \times V'$ in which $(x, x') \sim (y, y')$ if $(x, y) \in E$ and $(x', y') \in E'$. It follows easily from the definition that $G \times G'$ admits homomorphisms to G and G'. Therefore

$$\chi(G \times G') \leq \min\{\chi(G), \chi(G')\}$$

and

$$\chi_c(G \times G') \leq \min\{\chi_c(G), \chi_c(G')\}.$$

Equality is conjectured to hold in both inequalities.

Conjecture 5.1 ([Hed66]). *For any positive integer n, if $\chi(G) = \chi(G') = n$, then $\chi(G \times G') = n$.*

Conjecture 5.2 ([Zhu92]). *For any rational $r \geq 2$, if $\chi_c(G) = \chi_c(G') = r$, then $\chi_c(G \times G') = r$.*

Conjecture 5.2 is stronger than Conjecture 5.1. Both conjectures remain open, although the conjectures, especially Conjecture 5.1, have been studied extensively (see [Sau01, Zhu98]).

It is easy to see that Conjecture 5.1 is true if $n \leq 3$. The conjecture is also confirmed for $n = 4$ [ES85]. But this case is already quite difficult. For $n \geq 5$, Conjecture 5.1 remains open. Some other special cases of Conjecture 5.1 are proved: If $\chi(G) = \chi(G') = n$ and every vertex of G is contained in a clique of size $n - 1$, then $\chi(G \times G') = n$ [BEL76]. If $\chi(G) = \chi(G') = n$ and every pair of edges of G is connected by an edge of G, then $\chi(G \times G') = n$ [Tur83]. If $\chi(G) = \chi(G') = n$ and G is obtained from K_n by a series of Hajós sums and at most one contraction of non-adjacent vertices, then $\chi(G \times G') = n$ [SZ92]. If $\chi(G) = \chi(G') = n$, each of G and G' are connected, and $\omega(G) \geq n - 1$ and $\omega(G') \geq n - 1$, then $\chi(G \times G') = n$ [DSW85, Wel84]. More generally, Larose and Tardif [LT00] proved the following result: If K is vertex transitive and projective, then whenever K is the retract of the product of two connected graphs, it is a retract of a factor (see Section 10 for the definition of projective graphs). El-Zahar and Sauer [ES85] proposed the following conjecture which is stronger than Conjecture 5.1: *If G, G' are connected graphs with $\chi(G) = \chi(G') = n$ and H and H' are $(n-1)$-chromatic subgraphs of G and G', respectively, then the subgraph of $G \times G'$ induced by $(G \times H') \cup (H \times G')$ has chromatic number n.* This conjecture is disproved in [TZ02b]. Generally speaking, there is no substantial positive results concerning Conjecture 5.1 in the past twenty years. Even the following question is still open: Let $f(n) = \min\{\chi(G \times G') : \chi(G) = \chi(G') = n\}$. Is it true that $\lim_{n \to \infty} f(n) = \infty$? If Conjecture 5.1 is true, then we would have $f(n) = n$. It is proved by Poljak and Rödl [PR81] that $f(n)$ is either unbounded or is bounded by 16. This is strengthened in [Pol92] (see also [Zhu98]) where it is proved that $f(n)$ is either unbounded or bounded by 9.

A graph K is called *multiplicative* if $G \not\to K$ and $H \not\to K$ implies that $G \times H \not\to K$. Conjecture 5.1 says that each complete graph K_n is multiplicative. Conjecture 5.2 says that the circular complete graphs $K_{p/q}$ are multiplicative. As mentioned above, K_1, K_2, K_3 are the only complete graphs that are known to be multiplicative. However, there are some other graphs that are known to be multiplicative. The argument of El-Zahar and Sauer [ES85] is generalized by Häggkvist, Hell, Miller and Neumann Lara [HHMN88] to prove that odd cycles are multiplicative. Recently, Tardif [Tar05a] found an interesting graph operation that constructs new multiplicative graphs from old ones. This leads to a breakthrough in the study of Conjecture 5.2.

Suppose G is a graph. Then $\mathbf{P}_3^{-1}(G)$ is the graph defined as follows. For two subsets A, B of the vertex set of G, we write $A \bowtie B$ if every vertex of A is joined to every vertex of B. The graph $\mathbf{P}_3^{-1}(G)$ has vertex set $V(\mathbf{P}_3^{-1}(G)) = \{(u, A) : u \in V(G), \emptyset \neq A \subseteq N(u)\}$. Two vertices (u, A) and (v, B) are adjacent in $\mathbf{P}_3^{-1}(G)$ if $u \in B$, $v \in A$ and $A \bowtie B$.

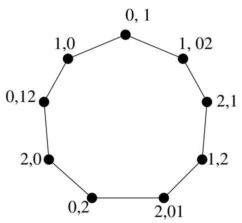


Fig. 3. The graph $\mathbf{P}_3^{-1}(K_3)$

The graph operation $\mathbf{P}_3^{-1}$ is the right inverse of a simpler graph operation: $\mathbf{P}_3$. Given a graph G, $\mathbf{P}_3(G)$ has vertex set $V(\mathbf{P}_3(G)) = V(G)$ in which xy is an edge if and only if there are vertices u, v such that xu, uv, vy are edges of G. I.e., G has an x-y-walk of length 3. Observe that if xy is an edge of G, then (x, y, x, y) is an x-y-walk of length 3, so xy is an edge of $\mathbf{P}_3(G)$. Thus G is a subgraph of $\mathbf{P}_3(G)$. If G has a triangle induced by x, y, z, then (x, y, z, x) is an x-x-walk of length 3. So $\mathbf{P}_3(G)$ has a loop. Conversely, $\mathbf{P}_3(G)$ has a loop only if G contains a triangle. Since we are only interested in loopless graphs, we shall apply the operator $\mathbf{P}_3$ to triangle free graphs only.

Lemma 5.3. *For any graph G, $\mathbf{P}_3(\mathbf{P}_3^{-1}(G))$ and G are homomorphically equivalent.*

However, in general, $\mathbf{P}_3^{-1}\mathbf{P}_3(G)$ and G are quite different. For example, $\mathbf{P}_3^{-1}\mathbf{P}_3(C_5) = \mathbf{P}_3^{-1}(K_5)$ is not homomorphically equivalent to C_5. Indeed, $\mathbf{P}_3^{-1}(K_5)$ is 5-chromatic, but C_5 is 3-chromatic. Nevertheless, we have the following lemma.

Lemma 5.4. *For any graphs G, H,*

1. $\mathbf{P}_3(G \times H) = \mathbf{P}_3(G) \times \mathbf{P}_3(H)$.
2. $\mathbf{P}_3(G) \to H$ *if and only if* $G \to \mathbf{P}_3^{-1}(H)$.
3. $G \to H$ *if and only if* $\mathbf{P}_3^{-1}(G) \to \mathbf{P}_3^{-1}(H)$.
4. $\mathbf{P}_3^{-1}(G \times H)$ *and* $\mathbf{P}_3^{-1}(G) \times \mathbf{P}_3^{-1}(H)$ *are homomorphically equivalent.*

Using Lemmas 5.4 and 5.4, one can easily prove the following:

Theorem 5.5. *A graph K is multiplicative if and only if $\mathbf{P}_3^{-1}(K)$ is multiplicative.*

Now we have a graph operator that might be used to construct new multiplicative graphs from old ones. But, unfortunately, applying the operator $\mathbf{P}_3^{-1}$ to the presently known multiplicative graphs (i.e., K_1, K_2 and odd cycles) does not yield new multiplicative graphs. It is easy to verify that $\mathbf{P}_3^{-1}(K_1)$ is empty, $\mathbf{P}_3^{-1}(K_2) = K_2$, and for any cycle C_n, $\mathbf{P}_3^{-1}(C_n) = C_{3n}$. At this step, it seems that this approach is leading to a dead end. "*Blocked by mountains and waters, from left to right. Where can I go? No road is in sight. Wait! There lies a village, where willows are dark, flowers are bright*". Instead of applying the operator $\mathbf{P}_3^{-1}$ to graphs that are known to be multiplicative, we apply this operator to graphs G which are not known to be multiplicative. If the resulting graph $\mathbf{P}_3^{-1}(G)$ is multiplicative, then Theorem 5.5 implies that G itself is multiplicative. To which graphs should we apply the operator $\mathbf{P}_3^{-1}$? Conjecture 5.2 provides a natural family of candidates: the circular complete graphs.

Theorem 5.6. *If $p/q < 12/5$, then $\mathbf{P}_3^{-1}(K_{p/(3q-p)})$ and $K_{p/q}$ are homomorphically equivalent.*

Corollary 5.7. *If $p/q < 12/5$, then $K_{p/q}$ is multiplicative if and only if $K_{p/(3q-p)}$ is multiplicative.*

Since for every $k \geq 3$, $K_{(2k+1)/k}$ is multiplicative and $(2k+1)/k < 12/5$, we can apply Corollary 5.7 and conclude that $K_{(2k+1)/(3k-(2k+1))} = K_{(2k+1)/(k-1)}$ is multiplicative. If $k \geq 9$, then $(2k+1)/(k-1) < 12/5$ and we can apply Corollary 5.7 again, and conclude that $K_{(2k+1)/(k-4)}$ is multiplicative. Continue the process, we have the following corollary:

Corollary 5.8. *If $k \geq 1$, $0 \leq i \leq \log_3(k)$ and $d = k - (3^i - 1)/2 > \frac{2k+1}{4}$, then $K_{(2k+1)/d}$ is multiplicative.*

It can be verified that the set $P = \{\frac{2k+1}{k-(3^i-1)/2} : k \geq 1, 0 \leq i \leq \log_3(k)\}$ is dense in the interval $[2, 4]$. Therefore, we have the following result:

Theorem 5.9 ([Tar05a]). *For any rational $2 \leq p/q < 4$, $K_{p/q}$ is multiplicative.*

The approach above can be interpreted in a different way: Although in general, $\mathbf{P}_3^{-1}$ is not a left inverse of $\mathbf{P}_3(G)$, but for some graphs G it may happen that $\mathbf{P}_3^{-1}\mathbf{P}_3(G)$ is homomorphically equivalent to G. Indeed, since $\mathbf{P}_3^{-1}$ is the right inverse of $\mathbf{P}_3$, for any graph G, $\mathbf{P}_3^{-1}\mathbf{P}_3(\mathbf{P}_3^{-1}(G))$ is homomorphically equivalent to $\mathbf{P}_3^{-1}(G)$. This means that if restricted to the family of graphs $\mathcal{H} = \{\mathbf{P}_3^{-1}(G) : G \text{ is a graph}\}$, $\mathbf{P}_3^{-1}$ is also a left inverse of $\mathbf{P}_3(G)$. For such graphs G, if G is multiplicative, then Theorem 5.5 implies that $\mathbf{P}_3(G)$ is multiplicative. Theorem 5.6 shows that if $p/q < 12/5$, then $K_{p/q} \in \mathcal{H}$. So if $p/q < 12/5$, and $K_{p/q}$ is multiplicative, then $\mathbf{P}_3(K_{p/q}) = K_{p/(3q-p)}$ is also multiplicative.

For a digraph G, let $\underline{G}$ be the underline graph G, i.e., $\underline{G}$ is obtained from G by omitting the orientation of the edges. It is known that $\chi(\underline{G \times H})$ could be strictly less than $\min\{\chi(\underline{G}), \chi(\underline{H})\}$. Indeed, even if G and H are tournaments, we can still have $\chi(\underline{G \times H}) < \min\{\chi(\underline{G}), \chi(\underline{H})\}$. Analogous to the undirected case, we define $g(n) = \min\{\chi(\underline{G \times H}) : G, H \text{ are digraphs with } \chi(\underline{G}) = \chi(\underline{H}) = n\}$. It is known that $g(n)$ is either bounded by 3 or unbounded [Pol92, Zhu98]. Recently, it is proved in [TW04] that $g(n)$ is bounded if and only if $f(n)$ (defined in the 5th paragraph of this section) is bounded. For the product of tournaments, Tardif [Tar04] defined $t(n) = \min\{\chi(\underline{G \times H}) : G \text{ and } H \text{ are } n\text{-tournaments}\}$, and proved that the sequence $t(n)/n$ tends to a limit λ, and $1/2 \leq \lambda \leq 2/3$.

There is also a conjecture analogue to Conjectures 5.1 and 5.2 for the fractional chromatic number of graphs [Zhu02c].

Conjecture 5.10. *If $\chi_f(G) \geq r$ and $\chi_f(H) \geq r$ then $\chi_f(G \times H) \geq r$.*

For $r \geq 2$, let $\phi(r) = \min\{\chi_f(G \times H) : \chi_f(G) = \chi_f(H) = r\}$. Although we do not know if $\phi(r) = r$, it is proved recently by Tardif [Tar05b] that $\phi(r) \geq r/4$. Tardif also considered the relation between the chromatic number of the product graph and the fractional chromatic number of the factor graphs. The following result is proved in [Tar01b].

Theorem 5.11. *If $\chi_f(G), \chi_f(H) \geq 2n$ then $\chi(G \times H) \geq n$.*

There are some other graph products whose circular chromatic number have been studied. For the Cartesian product $G \Box H$ (in which (x, y) is adjacent to (x', y') in $G \Box H$ if and only if either ($x = x'$ and $(y, y') \in E(H)$) or ($(x, x') \in E(G)$ and $y = y'$)), it is trivial that $\chi_c(G \Box H) = \max\{\chi_c(G), \chi_c(H)\}$. For the lexicographic product $G[H]$ (in which (x, y) is adjacent to (x', y') in $G[H]$ if $(x, x') \in E(G)$ or $x = x'$ and $(y, y') \in E(H)$), it is known [Zhu92] that $\chi_f(G)\chi(H) \leq \chi_c(G[H]) \leq \chi_c(G)\chi(H)$, and both the upper and lower bounds are sharp. Recently, the circular chromatic number and chromatic number of the Cartesian sum (also known as the very strong product) of two graphs is studied in [LZ05]. Suppose $G = (V, E)$ and $H = (V', E')$. The Cartesian sum $G \oplus H$ of G and H has vertex set $V \times V'$, in which (x, y) is adjacent to (x', y') in $G \oplus H$ if $(x, x') \in E(G)$ or $(y, y') \in E(H)$. The following result is a sharp upper bound for the chromatic number of $G \oplus H$ in terms of the circular chromatic number of G and H.

Theorem 5.12. *For any graphs G, H,*

$$\max\{\lceil \chi_f(G)\chi(H)\rceil, \lceil \chi(G)\chi_f(H)\rceil\} \le \chi(G \oplus H) \le \max\{\lceil \chi_c(G)\chi(H)\rceil, \lceil \chi(G)\chi_c(H)\rceil\}.$$

If $\chi_f(G) = \chi_c(G)$ and $\chi_f(H) = \chi_c(H)$, then the lower bound and upper bound above coincide. For general graphs G and H, it is known that $\max\{\chi_f(G)\chi(H), \chi(G)\chi_f(H)\} \le \chi_c(G \oplus H)$, and it is conjectured in [LZ05] that for any graphs G, H, $\chi_c(G \oplus H) \le \max\{\chi_c(G)\chi(H), \chi(G)\chi_c(H)\}$. Some special cases of the conjecture are confirmed in [LZ05], however, the general case is open.

6 Kneser, Schrijver and Cone Graphs (Topological Method)

Given positive integers $n \ge 2k$, the *Kneser graph* $K(n,k)$ has vertices all the k-subsets of $[n] = \{1, 2, \dots, n\}$ and two vertices u, v are adjacent if u and v do not intersect (as subsets of $[n]$). A k-subset u of $[n]$ is called *stable* if, by viewing the elements of $[n]$ as cyclically ordered, u does not contain two consecutive elements. Namely if $i \in u$ then $i+1 \notin u$, and if $n \in u$, then $1 \notin u$. The *Schrijver graph* $S(n,k)$ is the subgraph of $K(n,k)$ induced by those vertices which are stable k-subsets of $[n]$. The chromatic number of the Kneser graph $K(n,k)$ is conjectured by Kneser [Kne55] to be equal to $n - 2k + 2$. The conjecture remained open for more than 20 years, before it is proved by Lovász [Lov78] by an application of a topological method. Then Schrijver [Sch78] defined the graphs $S(n,k)$ and proved that $\chi(S(n,k)) = \chi(K(n,k)) = n - 2k + 2$ and moreover $S(n,k)$ is χ-critical, i.e., removing any vertex of $S(n,k)$ decreases its chromatic number.

The circular chromatic number of Kneser graphs is first studied by Johnson, Holroyd and Stahl [JHS97]. They conjectured that $\chi_c(K(n,k)) = \chi(K(n,k)) = n - 2k + 2$, and proved that the conjecture holds if $2k+1 \le n \le 2k+2$ or $k = 2$. The circular chromatic number of Schrijver graphs is first studied by Lih and Liu [LL02]. It is proved in [LL02] that if $k = 2$ and $n \ne 5$, then $\chi_c(S(n,k)) = \chi(S(n,k)) = 2-2k+2$. Since $S(2k+1,k)$ is the odd cycle C_{2k+1}, $\chi_c(S(2k+1,k)) \ne \chi(S(2k+1,k))$ for $k \ge 2$. Lih and Liu asked if for each k there is an integer $n(k)$ such that if $n \ge n(k)$ then $\chi_c(S(n,k)) = \chi(S(n,k))$. Hajiabolhassan and this author [HZ03c] proved Theorem 6.1 below which answers this question in the affirmative, and also provides strong support for the conjecture of Johnson, Holroyd and Stahl.

Theorem 6.1. *If $n \ge 2k^2(k-1)$, then $\chi_c(K(n,k)) = \chi(K(n,k)) = n-2k+2$. Moreover, for any integer k, there is an integer $n(k)$ such that if $n \ge n(k)$, then $\chi_c(S(n,k)) = \chi(S(n,k)) = n - 2k + 2$.*

Recently, Meunier [Meu05] and Simonyi and Tardos [ST04], used the topological tools in the study of the circular chromatic numbers of Kneser graphs and Schrijver graphs. Here we give a sketch of this approach.

Suppose $G = (V, E)$ is a graph. The *box complex* $B_0(G)$ of G is a simplicial complex with vertex set $V \times \{1, 2\}$, in which $(S \times \{1\}) \cup (T \times \{2\})$ is a simplex if $S \cap T = \emptyset$ and every vertex of S is adjacent to every vertex of T. We denote the set $(S \times \{1\}) \cup (T \times \{2\})$ by $S \uplus T$. For simplicity, the vertices $(x, 1)$ and $(x, 2)$ are denoted by $+x$ and $-x$, respectively.

As an example, we consider the box complex of complete graphs. Assume the vertex set of K_n is $\{v_1, v_2, \dots, v_n\}$. Then $B_0(K_1)$ consists of two 0-simplices (i.e., two sets of a single vertex), $\{+v_1\}, \{-v_1\}$. The complex $B_0(K_2)$ consists of four 0-simplices $\{+v_1\}, \{+v_2\}, \{-v_1\}, \{-v_2\}$ and four 1-simplices $\{+v_1, +v_2\}$, $\{-v_1, -v_2\}$, $\{+v_1, -v_2\}$, $\{+v_2, -v_1\}$. The complex $B_0(K_3)$ has six 0-simplices, twelve 1-simplices and eight 2-simplices, as shown in Figure 4.

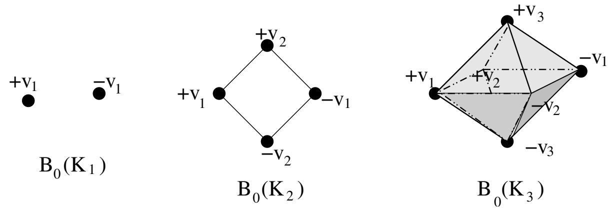


Fig. 4. The complexes $B_0(K_1)$, $B_0(K_2)$ and $B_0(K_3)$

In general, for $n \geq 2$, the complex $B_0(K_n)$ is obtained from $B_0(K_{n-1})$ by adding two new vertices $+v_n, -v_n$ and for each simplex σ of $B_0(K_{n-1})$, add the simplex $\sigma \cup \{+v_n\}$ and the simplex $\sigma \cup \{-v_n\}$. This means that $B_0(K_n)$ is the suspension of $(B_0(K_{n-1}))$. As the suspension of S^d is homeomorphic to S^{d+1}, and $B_0(K_1) = S^0$, we conclude that $B_0(K_n)$ is homeomorphic to S^{n-1} for all $n \geq 1$.

We shall apply Borsuk-Ulam Theorem in proving the lower bounds for the chromatic number and the circular chromatic number of Kneser graphs and Schriver graphs. For a topological space X, a Z_2*-action* on X is a homeomorphism $\nu : X \to X$ such that for any $x \in X$, $\nu^2(x) = x$. A Z_2-action is free if $\nu(x) \neq x$ for all x. A natural Z_2-action on the sphere S^d is the antipodal map $\nu(x) = -x$. The pair (X, ν) is called a Z_2-space. Suppose (x, ν) and (Y, μ) are Z_2-spaces. A continuous map $f : X \to Y$ is called a Z_2-map if $f(\nu(x)) = \mu(f(x))$ for all $x \in X$. We write $(X, \mu) \to (Y, \mu)$ if there is a Z_2-map from (X, ν) to (Y, μ). When the Z_2-action ν is clear from the context (like the anitpodal map on the spheres or $\mathbb{R}^n$), we simply write X instead of (X, ν). The following is one of the many equivalent formulations of Borsuk-Ulam Theorem.

Theorem 6.2. *For $d \geq 0$, there is no Z_2-map from S^d to S^{d-1}.*

For a graph G, the box complex of G has Z_2-action ν defined as $\nu(S \uplus T) = T \uplus S$. (Note that this definition defines ν on the vertice of the simplicial complex $B_0(G)$, and this definition extends to a topological realization of the simplicial complex by convex combination.) The link between colouring of graphs and box complexes of graphs is reflected in the following lemma, whose proof is straightforward.

Lemma 6.3. *Suppose G, H are graphs and ϕ is a homomorphism from G to H. Then the mapping ϕ^* defined as $\phi^*(S \uplus T) = \phi(S) \uplus \phi(T)$ is a Z_2-map from $B_0(G)$ to $B_0(H)$.*

In particular, if $\chi(G) \leq k$, then there is a homomorphism of G to K_k. By Lemma 6.3, this implies that $B_0(G) \to B_0(K_k) \approx S^{k-1}$. Therefore by Borsuk-Ulam Theorem, there is no Z_2-map from S^k to $B_0(G)$. In other words, we have

Corollary 6.4. *If there is a Z_2-map from S^k to $B_0(G)$, then $\chi(G) > k$.*

A graph G is called *topologically t-chromatic* if there is a Z_2-map from S^{t-1} to $B_0(G)$. As $S^{n-2k+1} \to B_0(S(n,k))$, so $\chi(K(n,k)) \geq \chi(S(n,k)) \geq n-2k+2$. The proof of the fact that $S^{n-2k+1} \to B_0(S(n,k))$ is not trivial, and it can be found in [ST04] and in the references given there. Our interest is why this gives a lower bound for the circular chromatic number of graphs.

Although there is no Z_2-map from S^d to S^{d-1}, if we delete a pair of antipodal points from S^d, then there is a Z_2-map from the resulting space to S^{d-1}. Indeed, let $u = (0, \dots, 0, 1)$ and $u' = (0, \dots, 0, -1)$ be the north pole and the south pole of S^d. Then $f : S^d \setminus \{u, u'\} \to S^{d-1}$ defined as $f((x_1 x_2, \dots, x_{d+1})) = \frac{1}{\sqrt{1-x_{d+1}^2}}(x_1, x_2, \dots, x_d)$ is a Z_2-map.

Theorem 6.5. *If a graph G is topologically k-chromatic and k is even, then $\chi_c(G) \geq k$.*

Proof. Assume $S^{k-1} \to B_0(G)$ and $k = 2t$ is odd. To prove that $\chi_c(G) \geq 2t$, by Theorem 3.4, it suffices to show that for any $2t$-colouring f of G, there is a cycle $C = (x_0, x_1, \dots, x_{2t-1})$ such that $f(x_{i+1}) = [f(x_i) + 1]_{2t}$ for $i = 0, 1, \dots, 2t-1$, where the summation in the indices are modulo $2t$.

Let f be a $2t$-colouring of G, which is viewed as a homomorphism from G to K_{2t}. Consider the induced Z_2-map $f^* : B_0(G) \mapsto B_0(K_{2t})$, which is a simplicial map. If f^* is not onto, then there is a simplex σ of $B_0(K_{2t})$ which is not the image of any simplex of $B_0(G)$ Thus f^* is a Z_2-map from $B_0(G)$ to $B_0(K_{2t}) \setminus \{\sigma, -\sigma\}$.

Since the deletion of a pair of antipodal points from S^{2t-1} results in a space which admits a Z_2-map to S^{2t-2}, and since $B_0(K_{2t})$ is homeomorphic to S^{2t-1}, we have $B_0(K_{2t}) \setminus \{\sigma, -\sigma\} \to S^{2t-2}$. This implies that

$$S^{2t-1} \to B_0(G) \to B_0(K_{2t}) \setminus \{\sigma, -\sigma\} \to S^{2t-2},$$

in contrary to Borsuk-Ulam Theorem. Therefore f^* must be onto. In particular, there is a simplex $S \uplus T$ of $B_0(G)$ which is mapped to $A \uplus A'$, where $A = \{1, 3, 5, \ldots, 2t-1\}$ and $A' = \{2, 4, \ldots, 2t\}$. So S contains vertices $v_1, v_3, \ldots, v_{2t-1}$ and T contains vertices $v_2, v_4, \ldots, v_{2t}$ with $f(v_i) = i$ for all i. As every vertex of S is adjacent to every vertex of T, $(v_1, v_2, \ldots, v_{2t})$ is a cycle in $D_f(G)$. □

Since $S(n,k)$ and $K(n,k)$ are topologically $(n-2k+2)$-chromatic and $\chi(S(n,k)) = \chi(K(n,k)) = n-2k+2$, we have the following corollary.

Corollary 6.6. *If $\chi(S(n,k))$ is even, then $\chi_c(K(n,k)) = \chi_c(S(n,k)) = \chi(S(n,k)) = \chi(K(n,k))$.*

Cone graphs is another class of graphs where the lower bound on its chromatic number is due to a topological reason. Suppose $G = (V, E)$ is a graph and $m \geq 1$ is an integer. The *m-cone* $\Delta_m(G)$ of G has vertex set

$$(V \times \{0, 1, \ldots, m\}) \cup \{u\},$$

and edge set

$$\{(x,0)(y,0) : xy \in E\} \cup \{(x,i)(y,i+1) : xy \in E, i = 0, 1, \ldots, m-1\}$$
$$\cup \{(x, m-1)u : x \in V\}.$$

We shall refer the m-cone over G for any $m \geq 0$ as a *cone* over G. Note that $\Delta_0(G)$ is the graph obtained from G by adding a universal vertex u and $\Delta_1(G) = M(G)$ is the Mycielskian of G. The cone over G is also called the *generalized Mycielskian of* G. The vertex u is called the *root* of $\Delta_m(G)$. For $\vec{m} = (m_1, m_2, \ldots, m_t)$, $\Delta_{\vec{m}}(G) = \Delta_{m_1}(\Delta_{m_2}(\ldots \Delta_{m_t}(G)\ldots))$. Figure 6 depicts the graph $\Delta_2(C_5)$.

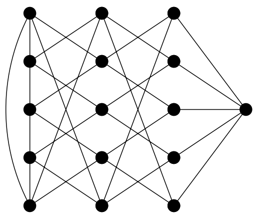


Fig. 5. The graph $\Delta_2(C_5)$

For $\vec{m} = (m_1, m_2, \ldots, m_t)$, $\Delta_{\vec{m}}(G) = \Delta_{m_1}(\Delta_{m_2}(\ldots \Delta_{m_t}(G)\ldots))$. It is easy to see that for any graph G, for any integer $m \geq 0$, $\chi(G) \leq \chi(\Delta_m(G)) \leq \chi(G)+1$. If $m \leq 1$, then $\chi(\Delta_m(G)) = \chi(G)+1$. For $m \geq 2$, it is possible that $\chi(\Delta_m(G)) = \chi(G)$. However, the following result is proved in [Sti85] (see also [GJS04]).

Theorem 6.7. *If* $\vec{m} = (m_1, m_2, \ldots, m_t)$*, then* $\chi(\Delta_{\vec{m}}(K_2)) = t + 2$*.*

The graph $\Delta_{\vec{m}}(K_2)$ is shown [Sti85] to have chromatic number $t+2$ also for a "topological reason". The topological reason considered in [Sti85] is different from the box complex argument, however, it is shown in [BK03] and [ST04] that that topological reason implies that $\Delta_{\vec{m}}(K_2)$ is topologically $(t+2)$-chromatic. Hence we have the following corollary [ST04].

Corollary 6.8. *If* $\vec{m} = (m_1, m_2, \ldots, m_t)$*, and* t *is even, then* $\chi_c(\Delta_{\vec{m}}(K_2)) = t + 2$*.*

On the other hand, the following result is proved by Lam, Lin, Gu and Song [LLGS03]:

Theorem 6.9. *If* n *is even and* $m \geq 0$ *is an integer, then* $\chi_c(\Delta_m(K_n)) = n + 1/t$*, where* $t = \lceil (2m+2)/n \rceil$*.*

Let $\vec{m'} = (m_2, m_3, \ldots, m_t)$ and $\vec{m} = (m_1, m_2, \ldots, m_t)$. Since $\Delta_{\vec{m'}}(K_2)$ admits a homomorphism to K_{t+1}, we conclude that $\Delta_{\vec{m}}(K_2)$ admits a homomorphism to $\Delta_{m_1}(K_{t+1})$. Therefore we have the following consequence.

Corollary 6.10. *If* t *is odd, then* $t + 1 < \chi_c(\Delta_{\vec{m}}(K_2)) \leq t + 1 + 1/s$*, where* $s = \lceil (2m_1 + 2)/(t+1) \rceil$*. In particular, for any* $\varepsilon > 0$*, there is an integer* n *such that if* t *is odd,* $\vec{m} = (m_1, m_2, \ldots, m_t)$ *and* $m_1 > n$*, then* $t + 1 < \chi_c(\Delta\vec{m}(K_2)) < t + 1 + \varepsilon$*.*

This result means that the condition that t be even in Corollary 6.8 is essential. The same is true for the circular chromatic number of Schrijver graphs. The following result is proved by Simonyi and Tardos [ST04]:

Lemma 6.11. *For any integers* $t, m \geq 1$*, if* n *is large enough, and* $n - 2k + 2 = t$*, then* $S(n,k)$ *admits a homomorphism to* $\Delta_m(K_{t-1})$*.*

As a consequence, we have the following corollary.

Corollary 6.12. *For any integers* $t, m \geq 1$*, if* n *is large enough and* $2k = n - 2t + 1$*, then* $2t < \chi_c(S(n,k)) < 2t + \varepsilon$*.*

7 Circular Perfect Graphs

Suppose G is a graph. The *clique number* of G is $\omega(G) = \max\{k : K_k \preceq G\}$, and the *circular clique number* of G is $\omega_c(G) = \max\{p/q : K_{p/q} \preceq G\}$. By this definition, for any graph G, we have

$$\omega(G) \leq \omega_c(G) \leq \chi_c(G) \leq \chi(G), \ \omega(G) = \lfloor \omega_c(G) \rfloor, \text{ and } \chi(G) = \lceil \chi_c(G) \rceil.$$

A graph G is *perfect* if for every induced subgraph H of G we have $\chi(H) = \omega(H)$. A graph G is called *circular perfect* if for every induced subgraph H

of G we have $\chi_c(H) = \omega_c(H)$. Circular perfect graphs are first defined in [Zhu05b], which was written in 2000, but published much later than some other papers on this subject.

As $\omega_c(H)$ and $\chi_c(H)$ are sandwiched between $\omega(H)$ and $\chi(H)$, every perfect graph is circular perfect. On the other hand, it is proved in [Zhu05b] that circular complete graphs are circular perfect. In particular, the odd cycles and their complements are circular perfect. So the family of circular perfect graphs is strictly larger than the family of perfect graphs. Some necessary conditions and sufficient conditions for a graph to be circular perfect are proved in [Zhu04, Zhu05b].

Theorem 7.1. *If G is a circular perfect graph, then for any vertex x of G, $N(x)$ induces a perfect subgraph of G.*

Theorem 7.2. *Suppose G is a graph such that for every vertex x of G, $N[x]$ induces a perfect graph and $G - N[x]$ is a bipartite graph with no induced path on* 5 *vertices. Then G is circular perfect.*

Observe that if $p/q \geq 3$, then $K_{p/q}$ satisfies the conditions of Theorem 7.2. So the circular perfectness of such a circular complete graph $K_{p/q}$ follows from Theorem 7.2. However, if $p/q < 3$, even the circular complete graph $K_{p/q}$ may not satisfy the condition of Theorem 7.2. If p/q is close to 2, the subgraph of $K_{p/q}$ induced by $K_{p/q} - N[x]$ may have a long induced path. By a close examination, one can extract a property that is shared by the induced paths contained in $K_{p/q} - N[x]$. Namely, none of these paths is badly linked with respect to x, where a 'badly linked path' is defined as follows:

Given an induced path $P_n = (p_0, p_1, \ldots, p_n)$ of $G - N[x]$, we say P_n is badly-linked *with respect to x if one of the following holds:*

1. *There are three indices $i < j < k$ of the same parity such that $N(p_i) \cap N(x) \not\subseteq N(p_j)$ and $N(p_k) \cap N(x) \not\subseteq N(p_j)$.*
2. *There are three indices $i < j < k$ of the same parity such that $N(p_j) \cap N(x) \not\subseteq N(p_i)$ and $N(p_j) \cap N(x) \not\subseteq N(p_k)$.*
3. *There are two even indices $i < j$ and two odd indices $i' < j'$ such that $N(p_i) \cap N(x) \not\subseteq N(p_j)$ and $N(p_{i'}) \cap N(x) \not\subseteq N(p_{j'})$.*

It turns out that the non-existence of badly linked path with respect to x in $G - N[x]$ is indeed crucial for the circular perfectness of a graph.

Theorem 7.3. *Suppose G is a triangle free graph such that for every vertex x of G, $G - N[x]$ is a bipartite graph with no induced C_n for $n \geq 6$, and no badly linked induced path with respect to x. Then G is circular perfect.*

A graph G is a convex round graph if its vertices can be cyclically ordered in such a way that the neighbourhood of each vertex is a 'consecutive segment' of the vertex set. Convex round graphs, which are shown to be circular perfect in [BH02], can be shown to satisfy the conditions of Theorem 7.2 or Theorem 7.3.

Theorems 7.2 and 7.3 are crucial in the proof of an analogue of Hajós Theorem for circular chromatic number. Hajós Theorem says that for any positive integer n, the family of graphs G with $\chi(G) \geq n$ is the minimal family of graphs which contains K_n and is closed under three operations: (1) taking the Hajós sum of two graphs, (2) identifying non-adjacent vertices, and (3) adding vertices and edges. The question of finding an analogue of Hajós Theorem for circular chromatic number is asked in [Zhu99a]. To find an analogue of Hajós Theorem for circular chromatic number, it amounts to find a set of graph operations that do not decrease the circular chromatic number, and can be used to construct all graphs G with $\chi_c(G) \geq p/q$, by starting from the single circular complete graph $K_{p/q}$. The operation of identifying non-adjacent vertices and the operation of adding vertices and edges do not decrease the circular chromatic number. However, taking the Hajós sum of two graphs with circular chromatic number p/q may results in a graph with circular chromatic number less than p/q. For example, the Hajós sum of two copies of K_3 is a C_5, where K_3 has circular chromatic number 3, and C_5 has circular chromatic number 5/2. So we need to find some other operations to replace the Hajós sum. It is proved in [Zhu01b], by using Theorem 7.2, that if $p/q \geq 3$, there is a set of three graph operations that can be used to replace the Hajós sum. The three graph operations are as follows:

The circular Hajós sum. Take $2k+1$ ($k \geq 1$) graphs $G_0, G_1, G_2, \ldots, G_{2k}$, with $e_i = x_i y_i$ be an edge of G_i. Remove the edges e_i for $i = 0, 1, 2, \ldots, 2k$. Identify all the x_i's into a single vertex x. Adding edges $y_0 y_1$, $y_1 y_2, \ldots$, $y_{2k-1} y_{2k}$, $y_{2k} y_0$. The resulting graph G is the *circular Hajós sum* of G_0, G_1, $G_2, \ldots, G_{2k}$.

The wheel join. Take n graphs $G_0, G_1, G_2, \ldots, G_{n-1}$ ($n \geq 4$), with $e_i = x_i y_i$ be an edge of G_i. Remove the edges e_i, and identify y_i with x_{i+1}, for $i = 0, 1, 2, \ldots, n-1$. Add edges to connect x_i and x_{i+2} for $i = 0, 1, 2, \ldots$, $n-1$, and add a vertex u and connect u to x_i for $i = 0y, 1, 2, \ldots, n-1$. (The additions in the indices are modulo n.) The resulting graph is the *wheel join* of $G_0, G_1, G_2, \ldots, G_{n-1}$.

The pentagon join. Take graphs $G_1, G_2, \ldots, G_7$, with $e_i = x_i y_i$ be an edge of G_i. Remove the edges e_i, and identify y_1, y_2, y_3, y_4, y_5 into a single vertex y. Identify x_6 with x_2, y_6 with x_5, x_7 with x_1, y_7 with x_4. Add edges $x_1 x_2, x_2 x_3, x_3 x_4, x_4 x_5$. The resulting graph is the *pentagon join* of $G_1, G_2, \ldots, G_7$.

Figure 6 is an illustration of these operations, where each shaded area represents a $G_i - e_i$.

It is not difficult to show that the three operations defined above do not decrease the circular chromatic number, i.e., if each G_i has $\chi_c(G_i) \geq r$, then the circular Hajós sum of the G_i's, the wheel join of the G_i's and the pentagon join of the G_i's all have circular chromatic number at least r (for the pentagon join, we need to assume that $r \geq 3$). However, it is non-trivial to show that these operations (together with identifying non-adjacent vertices and adding

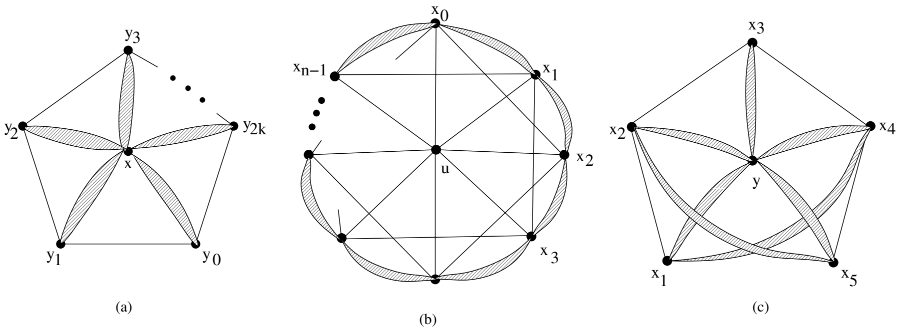


Fig. 6. (a) The circular Hajós sum (b) The wheel join (c) The pentagon join

vertices and edges) are enough to construct all graphs G with $\chi_c(G) \geq p/q$ (where $p/q \geq 3$), starting from the single graph $K_{p/q}$. This is done in [Zhu01b], where a crucial step relies on Theorem 7.2.

Theorem 7.4. *Suppose $p/q \geq 3$. Then the family of graphs G with $\chi_c(G) \geq p/q$ is the smallest family of graphs that contains $K_{p/q}$ and are closed under the operations of circular Hajós sum, wheel join, pentagon join, adding vertices and edges and identifying non-adjacent vertices.*

Theorem 7.4 does not apply to rationals $p/q < 3$. The corresponding result for $p/q < 3$ is proved in [Zhu03a], where four more graph operations are introduced. The descriptions of these four graph operations are technical. We omit the details.

An analogue of Hajós Theorem for circular chromatic number of edge weighted graphs is given in [Moh05]. The *strong Hajós sum* of two edge weighted graphs G_1 and G_2 with respect to the edges $u_1v_1 \in E(G_1)$ and $u_2v_2 \in E(G_2)$ is the graph obtained from the disjoint union of G_1 and G_2 by identifying u_1 and u_2, removing the edges u_1v_1 and u_2v_2, and adding an edge v_1v_2 with weight $a_{v_1v_2} = a_{u_1v_1} + a_{u_2v_2}$, where a_{uv} denotes the weight of edge uv. The following result is proved in [Moh05]:

Theorem 7.5. *For any positive real number r, any edge weighted graph G of circular chromatic number at least r contains a subgraph H (in the sense that the weight of an edge in H is at most the weight of that edge in G) which can be constructed by a sequence of identifications of nonadjacent vertices and strong Hajós sums, starting from edge weighted complete graphs of circular chromatic number r whose edge-weights satisfy the ε-triangle inequality: $a_{uv} \leq a_{uw} + a_{wv} + \varepsilon$.*

Unlike in Theorem 7.4, where three graph operations are used to in place of the Hajós sum, a single graph operation, the strong Hajós sum, is used in Theorem 7.5 in place of the original Hajós sum. As the compensation, Theorem 7.5 starts the construction not with a single graph, but with all complete edge weighted graphs (which is an infinite family of graphs) with circular chromatic number r whose edge-weights satisfy the ε-triangle inequality. Observe that the requirement that the edge-weights of the complete graphs satisfy the ε-triangle inequality is important, as any edge weighted graph can be made into a complete edge weighted graph by assigning 0 weight to the non-edges. On the other hand, it is unclear whether the conclusion remains true if the ε-triangle inequality is replaced by triangle inequality. Although the ε-triangle inequality is a small relaxation of the triangle inequality, however, in some sense, the strong Hajós sum allows the ε's to be accumulated.

A homomorphism from a graph G to a graph H is called an *H-colouring* of G. A generalization of Hajós theorem to H-colourings is obtained in [Neš99b], where a generalized Hajós sum is used to replace the Hajós sum to construct all non-H-colourable graphs, starting from a finite set of non-H-colourable graphs.

Parallel to the research on perfect graphs, one naturally wonders if there is an appealing circular perfect graph conjecture. A graph G is *circular imperfect* if it is not circular perfect. A graph G is *minimal circular imperfect* if G is circular imperfect but every proper induced subgraph of G is circular perfect. Can we characterize the family of all minimal circular imperfect graphs? Presently, we do not have a 'circular perfect graph conjecture'. A few families of graphs are shown to be minimal circular imperfect [PWZ05, PZ05, Xu05a, Xu05b, PZ06], which suggests that a characterization of all minimal circular imperfect graphs might be difficult.

Theorem 7.6. *The following graphs are minimal circular imperfect:*

1. *The complement of $K_{(3k+1)/3}$, i.e., the graph with vertex set $\{0, 1, \ldots, 3k\}$ in which ij is an edge if $j = i+1$ or $i+2$ (addition modulo $3k+1$).*
2. *The graph obtained from the odd wheel W_{2q+1} by subdividing each edge on the outer cycle into a path of length $2l-1$, where q, l are positive integers with $q + l \geq 3$.*
3. *The graph obtained from $K_{(2q+1)/2}$ by adding a universal vertex, and the graph obtained from $K_{(2q+1)/q}$ by adding a universal vertex.*

It was asked in [PWZ05] whether every minimal circular imperfect graph G has $\min\{\alpha(G), \omega(G)\} \leq 3$. The answer is negative. It is shown in [PZ06] that there are minimal circular imperfect graphs G with $\min\{\alpha(G), \omega(G)\}$ arbitrarily large.

Circular perfectness is not closed under complement. We call a graph G *strongly circular perfect* if both G and its complement are circular perfect. Perfect graphs are strongly circular perfect, odd cycles and their complements are strongly circular perfect. Strongly circular perfect graphs are studied in [CPW05, CPW05a, Yan05]. The following theorem follows from results in [Yan05].

Theorem 7.7. *If G is strongly circular perfect, then for any induced subgraph H of G, $\chi_c(H)$ is either an integer, or equal to $2 + 1/k$ or $k + 1/2$.*

Triangle-free strongly circular perfect graphs are characterized in [CPW05]. An *interlaced odd hole* is a graph obtained from an odd cycle C_{2k+1} as follows: Selecting two subsets A, B of $V(C_{2k+1})$ such that $A \cap B = \emptyset$ and B is an independent set. For each vertex v of B, replace v by a set S_v with $|S_v| \geq 2$ (so each vertex $x \in S_v$ is adjacent to the two neighbours of v in C_{2k+1}). For each vertex $v \in A$, add a set S_v of vertices so that each vertex of S_v is connected to v by an edge.

Theorem 7.8. *A triangle-free graph G is strongly circular perfect if and only if G is an interlaced odd hole.*

8 Colouring Subgraphs

Suppose G is a graph with $\chi_c(G) = p/q$. What can we say about the circular chromatic number of subgraphs of G? Some simple questions of this type remain open. For example, we do not know if it is true that any graph G with $\chi_c(G) = 8/3$ has a subgraph H with $\chi_c(H) = 5/2$. However, there are some progress in the study of the circular chromatic number of subgraphs.

The following result is proved in [HZ03a].

Theorem 8.1. *Suppose G is a graph and e is an edge of G. Then $\chi_c(G-e) \geq \lceil \chi_c(G) \rceil - 1$.*

Corollary 8.2. *If n is a positive integer and G is a graph with $\chi_c(G) > n$, then G has a subgraph H with $\chi_c(H) = n$.*

Corollary 8.2 provides an easy way to construct large girth graphs G with $\chi_c(G)$ equal to a given positive integer. One only needs to construct a graph G of large girth and with $\chi(G) > n$, then by removing some edges from G, one can obtain a graph G' (which still has large girth) with $\chi_c(G') = n$.

Corollary 8.3. *If $\chi(G) = n+1$ and $\chi(G-e) = n$ for an edge e of G, then $\chi_c(G-e) = n$.*

Corollary 8.4. *If $\chi(G) = n$ and G has a vertex x such that for every n-colouring f of G, $|f(N[x])| = n$, then $\chi_c(G) = n$.*

Corollary 8.4 follows from Corollary 8.3 by considering the graph G' obtained from G by adding a vertex x' and connecting x' to x and to each neighbour of x by an edge. The condition implies that $\chi(G') = \chi(G) + 1$, and hence $\chi_c(G) = \chi_c(G' - xx') = \chi(G') - 1 = \chi(G)$.

The question whether every graph G has a vertex x with $\chi_c(G-x) \geq \chi_c(G) - 1$ is asked in [Zhu01c]. This question is answered in negative in [Zhu04b].

Theorem 8.5. *There is an infinite family of graphs G such that $\chi_c(G) = 4$ and for any vertex x of G, $\chi_c(G-x) = 8/3$.*

Figure 7 is an example of a graph G with $\chi_c(G) = 4$ and $\chi_c(G - x) = 8/3$ for each vertex x. The integers besides the vertices show a $(8,3)$-colouring of $G - w$. Note that the graph G is not vertex transitive. The automorphism group of G has three orbits. To prove that $\chi_c(G-x) \leq 8/3$ for every vertex v, we need to find a $(8,3)$-colouring for $G - v$ and $G - u$ as well.

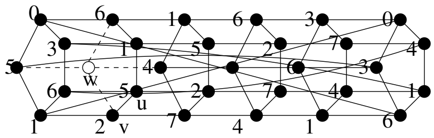


Fig. 7. A graph G with $\chi_c(G) = 4$ and $\chi_c(G - x) = 8/3$ for each vertex x

Although there is an infinite family of graphs G for which $\chi_c(G - x) < \chi_c(G) - 1$ for each vertex x, all these example graphs have similar structure. In particular, all the graphs have circular chromatic number 4. It is not clear if 4 is an exceptional integer. Properties of graphs G for which $\chi_c(G-x) < \chi_c(G)-1$ for each vertex x are studied in [Xu05c].

Question 8.6. Does there exist a graph G with $\chi_c(G) \neq 4$ and for every vertex x of G, $\chi_c(G - x) < \chi_c(G) - 1$?

Given a fraction p/q with $(p,q) = 1$. The fraction p'/q' with $p' < p$ and $pq' - p'q = 1$ is the fraction precedes p/q in the Farey sequence. It is known [BH90, Zhu92] that for any vertex x of the circular complete graph $K_{p/q}$, $\chi_c(K_{p/q} - x) = p'/q'$. The following question remains open.

Question 8.7. Suppose G is a graph with $\chi_c(G) = p/q$, where $(p,q) = 1$. Is it true that G has a subgraph H with $\chi_c(H) = p'/q'$?

Theorem 8.1 implies that if p/q is an integer, then the answer to Question 8.7 is positive. For some fractions p/q, say $p/q \neq 4$, it might be true that every graph G with $\chi_c(G) = p/q$ has an induced subgraph H with $\chi_c(H) = p'/q'$.

All known graphs having a vertex v with $\chi_c(G - v) \leq \chi_c(G) - 1$ satisfy $\chi_c(G) = \chi(G)$. This motivates the following question:

Question 8.8. Suppose G is a graph with $\chi_c(G) < \chi(G)$. Is it true that for any vertex x of G, $\chi_c(G - x) \geq \chi(G) - 1$?

9 K_n-minor Free Graphs

In [Vin88], Vince showed that for any rational $p/q \geq 2$, there is a graph G, namely $G = K_{p/q}$, which has circular chromatic number p/q. (The notation for the graph $K_{p/q}$ has been evolved from $G(q,p)$ [Vin88] to G^q_p [BH90], and hopefully settled down to $K_{p/q}$. The graph $K_{p/q}$ can also be defined as the complement of powers of cycles $\overline{C^q_p}$, and are known as *antiwebs* in the literature [She95, Wag03].) If we want to construct a graph with some special properties to have a given circular chromatic number, or to prove there is no such graph, the problem can be very difficult.

A graph H is a *minor* of a graph G if H can be obtained from a subgraph of G by contracting some edges. A graph G is called *H-minor free* if H is not a minor of G. The well-known Hadwiger's conjecture says that K_n-minor free graphs are $(n-1)$-colourable. If this conjecture is true, then every K_n-minor free graph has circular chromatic number at most $n-1$. It is natural to ask if every rational number $r \in [2, n-1]$ is the circular chromatic number of some K_n-minor free graphs. The answer is positive if $n \geq 5$ and negative for $n = 4$ [HZ00, LPZ03, Zhu00]. There are two key ideas in the construction of K_n-minor free graphs with given circular chromatic number. These ideas are useful elsewhere, for example, in the construction of graphs with given circular flow number and the construction of uniquely colourable graphs with given circular chromatic number. One of the ideas is to use Farey sequence that constructs the fractions between two consecutive integers in a special order. Given a non-negative integer n, the fractions in the interval $[n, n+1]$ can be partially ordered as follows: First we have two fractions $n/1$ and $(n+1)/1$. Suppose a/b and a'/b' are two fractions with $a'b - ab' = 1$, then we can construct a new fraction $p/q = (a+a')/(a+b')$. We call a/b and a'/b' the *lower parent* and the *upper parent* of p/q, respectively, and denoted by $a/b = p_l(p/q)$ and $a'/b' = p_u(p/q)$. Starting from $n/1$ and $(n+1)/1$, all the fractions $p/q \in [n, n+1]$ can be constructed in this way. If we want to construct, for a given $p/q \in [n, n+1]$, a K_{n+2}-minor free $\chi_c(G) = p/q$, we shall use induction, by first constructing K_{n+2}-minor free graphs H and H' with $\chi_c(H) = p_l(p/q)$ and $\chi_c(H') = p_u(p/q)$. Then graphs H and H' are used as building blocks in the construction of G.

The second idea is the labeling method. Suppose G is a graph and e^* is an edge of G. The pair (G, e^*) is called a *rooted graph* with *root edge* e^*. For convenience, we usually say G is a rooted graph with root edge e^*. Suppose G is a rooted graph with root edge $e^* = xy$ and $r \geq 2$ is a real number. Let D be an orientation of G. For convenience, the arc obtained by assigning an orientation to e^* is also denoted by e^*. A *rooted r-tension* of D is a tension f on D such that for each arc $e \neq e^*$, $1 \leq |f(e)| \leq r-1$. In other words, a rooted r-tension f is almost an r-tension except that there is no constraint on the value of f on the root arc.

Suppose G is a rooted graph with root edge e^* and $r \geq 2$ is a real number. Let D be an orientation of G. The *r-tension label set* of (G, e^*), denoted by $L^r_T(G, e^*)$, is defined as

$$L^r_T(G, e^*) = \{t \in [0, r), \exists \text{ a rooted } r\text{-tension } f \text{ of } D \text{ with } f(e^*) = t\}.$$

Although the definition of $L_T^r(G, e^*)$ needs to refer to an orientation of G, but any orientation defines the same label set.

We usually write $L_T^r(G)$ for $L_T^r(G, e^*)$. The root edges are usually clear from the context (or is irrelevant, such as in K_n or C_n). As an example, it follows from the definition that $L_T^r(C_2) = [1, r-1]_r$, where C_2 is the cycle with two edges.

The r-tension label set of a rooted graph G with root edge e^* contains information of possible r-tensions of $G - e^*$. For example, it is easy to see from the definition that

$$\begin{aligned}\chi_c(G) \leq r &\Leftrightarrow [1, r-1] \cap L_T^r(G) \neq \emptyset \\ \chi_c(G - e^*) \leq r &\Leftrightarrow L_T^r(G) \neq \emptyset \\ \chi_c(G/e^*) \leq 0 &\Leftrightarrow 0 \in L_T^r(G).\end{aligned}$$

In general, it is difficult to calculate the r-tension label set of a rooted graph. However, if a rooted graph G is constructed from some simple building blocks by some simple graph operations, then it might be easy to calculate $L_T^r(G)$. Here by 'simple graph operations', we mean the series join, the parallel join, the extended series joins and the diamond constructions. Among these, the easiest are series join and parallel join.

Suppose G and G' are two rooted graphs, with root edges $e = xy$ and $e' = x'y'$, respectively.

Series join. The series join of G and G', denoted by $S(G, G')$, is the rooted graph obtained from the disjoint union of G and G' by removing e and e', identifying y and x' into a single vertex z, and adding an edge $e'' = xy'$, where e'' is the root edge of G''.

Parallel join. The parallel join of G and G', denoted by $P(G, G')$, is the rooted graph obtained from the disjoint union of G and G' by removing e and e', identifying x and x' into a single vertex x'', identifying y and y' into a single vertex y'', and adding an edge $e'' = x''y''$, where e'' is the root edge of G''.

For two subsets A, B of $[0, r)$, let $A + B = \{[a + b]_r : a \in A, b \in B\}$. Then we have the following lemma:

Lemma 9.1. *For any rooted graphs G, G', we have*

$$\begin{aligned}L_T^r(S(G, G')) &= L_T^r(G) + L_T^r(G') \\ L_T^r(P(G, G')) &= L_T^r(G) \cap L_T^r(G').\end{aligned}$$

Now we ask the following question: Let $\mathcal{F}$ be a family of graphs (usually $\mathcal{F}$ consists of a single or very few small graphs). Let $r \geq 2$ be a rational. Can we construct a graph G with $\chi_c(G) = r$ by repeatedly applying the series join and parallel joins to a graphs in $\mathcal{F}$?

Instead of constructing a graph with a given circular chromatic number, we may construct a graph with a given r-tension label set. This seemingly more difficult task turns out to be easier. Indeed, we may completely forget about graphs, and simply construct their label sets. Formally, fix a real number r and an initial label sets $\mathcal{L}$ (of subsets of $[0, r)$). We define *constructible label sets* (with respect to r) as follows: *If $A \in \mathcal{L}$, then A is constructible. If A, B are constructible label sets, then $A \cap B$ and $A + B$ are constructible label sets.* Note that $A + B$ is taken as the sum two subsets of $[0, r)$.

It follows from Lemma 9.1 that if $\mathcal{L} = \{L_T^r(G) : G \in \mathcal{F}\}$, then a label set A is constructible if and only if there is a graph G which can be constructed by repeatedly applying the series join and the parallel join to graphs in $\mathcal{F}$, and for which we have $L_T^r(G) = A$.

Theorem 9.2. *Suppose $\mathcal{L}$ is an initial label set and $\mathcal{A}$ is the set of label sets constructible from $\mathcal{L}$. Suppose $p/q > 4$ and $p_l(p/q) < r < p_u(p/q)$. If $[n, r - n]_r, [1, r - 1]_r \in \mathcal{L}$, then*

$$[p - 1 - (q - 1)r, qr - p + 1]_r \in \mathcal{A}_r.$$

Observe that for any $r > 2$, $L_T^r(C_2) = [1, r - 1]_r$. Suppose H is a graph and e is an edge of H. It is not difficult to prove that if $\chi_c(H) = n + 1$ and $\chi_c(H - e) = n$, then for any $r \in (n, n + 1)$, $L_T^r(H) = [n, r - n]_r$. Thus as a consequence of Theorem 9.2, we have the following result:

Theorem 9.3. *Let $\mathcal{H}$ be a family of rooted graphs which contains C_2 and are closed under series joins and parallel joins. Let $n \geq 4$ be an integer. If there is a rooted graph $(H, e) \in \mathcal{H}$ such that $\chi_c(H) = n + 1$ and $\chi_c(H - e) = n$, then for every rational $r \in (n, n + 1)$, there is a rooted graph $G \in \mathcal{H}$ such that $\chi_c(G) = r$.*

As for $n \geq 4$, we have $\chi_c(K_{n+1}) = n + 1$ and $\chi_c(K_{n+1} - e) = n$ for any edge e, we have the following result, which is proved in [LPZ03], also answering a question in [Zhu01c]:

Theorem 9.4. *If $n \geq 5$, then for any rational number $r \in [n - 2, n - 1]$, there is a K_n-minor free graph G with $\chi_c(G) = r$.*

It is proved earlier in [Mos97] and [Zhu99b] that for any rational $r \in [2, 4]$, there is a planar graph (and hence a K_5-minor free graph) G with $\chi_c(G) = r$. Combined with the results above, we have

Corollary 9.5. *If $n \geq 5$, then for any rational $r \in [2, n - 1]$, there is a K_n-minor free graph G with $\chi_c(G) = r$.*

For K_4-minor free graphs, the following result is a combination of results in [HZ00] and [PZ04].

Theorem 9.6. *If G is a K_4-minor free graph, then either $\chi_c(G) = 3$ or $\chi_c(G) \leq 8/3$. Moreover, for any rational $r \in [2, 8/3] \cup \{3\}$, there is a K_4-minor free graph G with $\chi_c(G) = r$.*

In the construction of planar graphs with circular chromatic number r for $r \in [2,4]$, the series joins and parallel joins are not enough. Two other graph operations have been used: diamond construction and extended series joins. We refer readers to [LPZ03] for discussions about these two operations.

10 Graphs of Large Girth

A classical result of Erdős says that graphs of large girth can have large chromatic number. Erdős' result has been strengthened in different ways. One very strong result in this aspect is the following result of Müller [Mül79].

Theorem 10.1. *Let ℓ be an integer, let A be a finite set, and let $f_1, f_2, \ldots, f_t$ be distinct n-colourings of the elements of A. Then there exists a graph $G = (V, E)$ such that the following hold:*

- *A is a subset of V.*
- *Each n-colouring of G is an extension of some f_i, and for each $i = 1, 2, \ldots, t$, there exists a unique n-colouring g_i of G which is an extension of f_i.*
- *G has girth at least ℓ.*

If $t = 1$, then the graph G as in Theorem 10.1 is uniquely n-colourable, and hence has chromatic number n.

Theorem 10.1 can be generalized to circular colourings, and more generally to H-colourings. Two r-colourings f and g of a graph G are equivalent if there are constants $c \in [0, r)$ and $\tau \in \{-1, 1\}$ such that for any vertex x of G, $f(x) = [\tau g(x) + c]_r$. A graph G is called *uniquely r-colourable* if G has exactly one r-colouring, up to equivalence. Two H-colourings f, g of a graph G are equivalent if there is an automorphism σ of H such that $f = \sigma \circ g$. A graph G is called uniquely H-colourable if, up to equivalence, there is exactly one onto H-colouring of G. If $r = p/q \geq 2$ and $(p, q) = 1$, then a graph G is uniquely r-colourable if and only if G is uniquely $K_{p/q}$-colourable. It is easy to see that if a graph G is uniquely r-colourable, then $\chi_c(G) = r$. So the existence of uniquely r-colourable graphs with a certain property implies the existence of graphs G with $\chi_c(G) = r$ with that property. One question of interest is the existence of graphs of large girth with given circular chromatic number [AZ93, NZ01, NZ04, SZ96, Zhu96a, Zhu99d]. For any rational $r \geq 2$, for any integer g, does there exist a graph G of girth at least g and with $\chi_c(G) = r$? This leads to the study of the existence of uniquely r-colourable graphs of large girth. A graph G is a *core* if it does not admit a homomorphism to a proper subgraph. It is proved in [Zhu96a] that for any core graph H and for any integer g, there is a graph G of girth at least g such that G is uniquely H-colourable. This implies that for any $r \geq 2$ and for any integer g, there is a graph G of girth at least g and with $\chi_c(G) = r$. In [NZ04], the following question is considered: Let H be a graph. Given any integers t, g, does there exist a graph G of girth at least g which has exactly t H-colourings, up to equivalence? In case $t = 1$, the question asks whether there exists a uniquely H-colourable graph G of girth at least g. It turns our the answer is positive in most cases. Indeed, in most cases, we can ask for more: not only there are exactly t H-colourings, but these t H-colourings are extensions of any previously given t mappings from a subset A of $V(G)$ to $V(H)$. To state the result, we need one definition. A graph H is called *t-projective* if every homomorphism $f : H^t \to H$ is equivalent to a projection, i.e., there is a projection p and an automorphism σ of H such that $f = \sigma \circ p$. Here H^t is the categorical product of t copies of H, and a projection from H^t to H is a homomorphism $p : H^t \to H$ of the form $p(x_1, x_2, \ldots, x_t) = x_i$ for a fixed i. A graph H is called *projective* if H is t-projective for every positive integer t. Note that a projective graph is necessarily a core graph. Projectivity of graphs is first defined in [LT00, LT01], but studied by others in different contexts (see [ŁN04]). It is proved in [LT00, LT01] that a graph H is projective if and only if H is 2-projective. In [ŁN04], it is proved that most graphs are projective. The following result is proved in [NZ04]:

Theorem 10.2. *Let H be a projective graph with k vertices. Let ℓ be an integer, let A be a finite set, and let $f_1, f_2, \ldots, f_t$ be distinct mappings from A to $V(H)$. Then there exists a graph $G = (V, E)$ such that the following hold:*

- *A is a subset of V.*
- *For every $i = 1, 2, \ldots, t$, there exists a unique homomorphism $g_i : G \to H$ such that $g_i|_A = f_i$.*
- *For every homomorphism $f : G \to H$, there exists an index i, $1 \le i \le t$, and an automorphism σ of H such that $f = \sigma \circ g_i$.*
- *G has girth at least ℓ.*

It follows from some general results in [LT00, LT01, LT02] that the circular complete graphs are projective. A direct proof of this fact can also be found in [NZ01]. Thus we have the following corollary:

Corollary 10.3. *For any rational $r \ge 2$, for any positive integers t, ℓ, there is a graph G of girth at least ℓ such that G has exactly t r-colourings, up to equivalence. In particular, for any $r \ge 2$ and for any integer ℓ, there is a graph G of girth at least ℓ which is uniquely r-colourable and hence it has $\chi_c(G) = r$.*

The proof of Theorem 10.2 in [NZ04] uses the probabilistic method. A constructive proof of Corollary 10.3 for $r \ge 3$ is given in [NZ01], and a constructive proof of the corollary for $2 \le r \le 3$ is given in [PZ05].

In [EHK98], it is proved that there exist uniquely n-colourable graphs of large girth with bounded maximum degree. This result is generalized in [HZ04], where it is proved that for any rational p/q, for any integer g, there is a uniquely p/q-colourable graph G of girth g and with $\Delta(G) \le 5p^{13}$.

One interesting problem is the circular chromatic number of cubic graphs of large girth. The following question is due to Nešetřil [Neš99a]:

Question 10.4. Is it true that cubic graphs G of sufficiently large girth have $\chi_c(G) \le 5/2$?

The answer is negative if 5/2 is replaced by 7/3 [Hat05]. The following question is also open.

Is it true that cubic graphs G of sufficiently large girth have $\chi_c(G) \le r$ for some $r < 3$?

On the other hand, it is proved by Hatami and this author [JZ05] that cubic graphs G of girth at least 4 have $\chi_f(G) \le 3 - \frac{3}{64}$.

11 $\chi_c(G)$ vs. $\chi(G)$ and $\chi_f(G)$

It is known that for any graph G, we have $\chi_f(G) \le \chi_c(G) \le \chi(G)$. Graphs G with $\chi_c(G) = \chi(G)$ and graphs with $\chi_f(G) = \chi_c(G)$ are of special interest and have been studied in many papers. It is proved in [Gui93] that it is NP-hard to determine if a graph G satisfies $\chi_c(G) = \chi(G)$. In [Zhu01c], the author asked the question whether the problem is still NP-hard if $\chi(G)$ is known. This question is answered in [HT04].

Theorem 11.1. *The following decision problem is NP-hard:*
Instance: A graph G with $\chi(G) = n$;
Question: Is it true that $\chi_c(G) = \chi(G)$?

Some sufficient conditions are listed in [Zhu01c] under which a graph G has $\chi_c(G) = \chi(G)$. For example, if G is uniquely n-colourable, then $\chi_c(G) = \chi(G)$ [Zhu96a]; if the complement of G is disconnected, then $\chi_c(G) = \chi(G)$ [AZ93, Zhu92]. The following is a new sufficient condition found by Fan [Fan04].

Theorem 11.2. *If the complement of G is not Hamiltonian, then $\chi_c(G) = \chi(G)$.*

The proof of Theorem 11.2 is quite easy. Assume $\chi_c(G) = p/q$, where $(p, q) = 1$ and $q \ge 2$. We shall prove that the complement $\overline{G}$ of G has a Hamilton cycle. Let f be a (p, q)-colouring of G. Then for each i, $f^{-1}(i) \ne \emptyset$. Let X_i be an arbitrary ordering of the vertices of $f^{-1}(i)$, then the cyclic concatenation $X_0 X_1 \dots X_{p-1}$ of the X_i's is a Hamilton cycle of $\overline{G}$.

Theorem 11.2 is used in [Fan04] to study the circular chromatic number of iterated Mycielskian of complete graphs. It is known that $\chi(M(G)) = \chi(G) + 1$, but $\chi_c(M(G))$ could be strictly smaller than $\chi(G) + 1$. However, for any integer $k \ge 3$, $\chi_c(M(K_n)) = n + 1$ [CHZ99]. For an integer $t \ge 2$, the t-th iterated Mycielskian $M^t(G)$ of G is defined as $M^t(G) = M(M^{t-1}(G))$. The circular chromatic number of the iterated Mycielskian of complete graphs is studied in [CHZ99]. The following conjecture is proposed in [CHZ99].

Conjecture 11.3. *If $t \le n - 2$, then $\chi_c(M^t(K_n)) = \chi(M^t(K_n))$.*

As $M(G) = \Delta_1(G)$, Corollary 6.8 implies that $\chi_c(M^t(K_n)) = \chi(M^t(K_n))$ if $t+n$ is even. On the other hand, it is shown in [CHZ99] that if $t = n-1$, then $\chi_c(M^t(K_n)) \leq \chi(M^t(G)) - 1/2$. So the requirement that $t \leq n-2$ in Conjecture 11.3 is needed.

For $n = 3, 4$, Conjecture 11.3 is verified [CHZ99]. For $n = 3$, a stronger result is proved in [AHT03]: *If $n \geq 3$ and G uniquely n-colourable, then $\chi_c(M(G)) = \chi(M(G))$.* Theorem 11.2 provides a simple proof for the $n = 3$ case of Conjecture 11.3, as $\overline{M(K_n)}$ (for $n \geq 3$) is easily seen to be non-Hamiltonian. By using Theorem 11.2, it is also proved in [Fan04] that if G has n vertices and $\chi(G) \geq (n+3)/2$, then $\chi_c(M(G)) = \chi(M(G))$.

For $n \geq 5$, Conjecture 11.3 remains open. However, it is proved in [HZ03b] that if $n \geq 2^t + 2$, then $\chi_c(M^t(K_n)) = \chi(M^t(K_n))$. This result is improved in [Liu04], where it is proved that if $m \geq 2^{t-1} + 2t - 2$ and $t \geq 2$, then $\chi_c(M^t(K_m)) = \chi(M^t(K_m))$.

Graphs G with $\chi_f(G) = \chi_c(G)$ are called *star extremal*. Star extremal graphs have some nice properties. For example, if G, H are star extremal, then the upper and lower bounds in Theorem 5.12 coincide, and hence the chromatic number of $G \oplus H$ is determined. It is likely that it is NP-hard to determine if an arbitrary graph G is star extremal, but as far as this author knows, there is no proof yet. The study of star extremal graphs has been concentrated on circulant graphs and distance graphs. Some classes of circulant graphs and distance graphs are proved to have this property [CHZ98, CLZ99, GZ96, HC00, LLZ99, Lin03, LG04, LLS05, LZ99, WL04, Zhu96b, Zhu02a]. A useful tool in the study of the circular chromatic number of circulant graphs and distance graphs is the *regular colouring method*. Suppose n is a positive integer and D is a set of positive integers $i \leq n/2$. The circulant graph $G(n, D)$ has vertex set $\{0, 1, \ldots, n-1\}$ in which ij is an edge if $[i-j]_n \in D$ or $[j-i]_n \in D$. Given an integer k, let $\lambda_k(D) = \min\{|ki|_n : i \in D\}$, and let $\lambda(D) = \max\{\lambda_k(D) : k = 1, 2, \ldots, n\}$. It is known [GZ96] that $\chi_c(G(n, D)) \leq n/\lambda(D)$. In many cases, this bound is sharp. For example, if $G(n, D)$ has maximum degree 3, then $\chi_c(G(n, D)) = n/\lambda(D) = \chi_f(G(n, D))$ [GZ96]; if $D = \{i, i+1, i+2, \ldots, i'\}$ and $i' \geq 6i/5$, then $\chi_c(G(n, D)) = n/\lambda(D) = \chi_f(G(n, D))$ [LLZ99].

For a finite set D of positive integers, the distance graph $G(Z, D)$ has vertex set Z in which ij is an edge if $|i-j| \in D$. For any integers k, n, let $\lambda_k^n(D) = \min\{|ki|_n : i \in D\}$, and let $\lambda^n(D) = \max\{\lambda_k^n(D) : k = 1, 2, \ldots, n\}$. It is also known that $\chi_c(G(Z, D)) \leq n/\lambda^n(D)$ for any integer n. In many cases, this upper bound is also sharp for an appropriately chosen n.

12 Circular Choosability

A k-list assignment of a graph G is a mapping L which assigns to each vertex x a set $L(x)$ of k colours. A G is called k-choosable if for each k-list assignment L, there is a colouring f of G such that for each vertex x, $f(x) \in L(x)$. The *choosability* (or the *list chromatic number*) of G is the least k for which G is k-choosable. Choosability of graphs is an extensively studied graph parameter. The concept of choosability of graphs can be naturally extended to circular colouring. Suppose G is a graph and $p \geq 2q$ are positive integers. A *(p,q)-list assignment* L is a mapping which assigns to each vertex v of G a subset $L(v)$ of $\{0, 1, \dots, p-1\}$. An *L-(p,q)-colouring* of G is a (p,q)-colouring f of G such that for any vertex v, $f(v) \in L(v)$. A *list-size assignment* is a mapping $\ell : V \to [0, p/q]$. Given a list-size assignment ℓ, an ℓ-(p,q)-list assignment is a (p,q)-list assignment L such that for each vertex v, $|L(v)| \geq \ell(v)q$. A graph G is called *ℓ-(p,q)-choosable* if for any ℓ-(p,q)-list assignment L, G has an L-(p,q)-colouring.

The circular list chromatic number of a graph is defined through list-size assignments which are constant mappings. Suppose $t \geq 1$ is a real number and $p \geq 2q$ are positive integers. A *t-(p,q)-list assignment* is a (p,q)-list assignment L such that for every vertex v, $|L(v)| \geq tq$. We say G is *circular t-(p,q)-choosable* if for any t-(p,q)-list assignment L, G has an L-(p,q)-colouring. We say G is *circular t-choosable* if G is circular t-(p,q)-choosable for any positive integers $p \geq 2q$. The *circular list chromatic number* (or the *circular choosability*) of G is defined as

$$\chi_{c,l}(G) = \inf\{t : G \text{ is circular } t\text{-choosable}\}.$$

The circular list chromatic number of a graph can also be defined through circular r-colourings using colours from the circle S^r. A subset U of S^r is said to be *assignable* if it is the union of finitely many disjoint open arcs on S^r. The *length of an assignable set U*, denoted by $\text{length}(U)$, is the sum of the lengths of the open arcs of U. If $G = (V, E)$ is a graph, then a function L that assigns to each vertex v of G an assignable subset $L(v)$ of S^r is called an *circular list assignment* (with respect to r). If for each vertex v of G, $L(v)$ has length at least t, then L is called a *t-circular list assignment* (with respect to r). A *circular L-colouring* of G is a mapping c from V to S^r such that $c(v) \in L(v)$ for each vertex v of G and for every pair u, v of adjacent vertices of G, $|c(u) - c(v)|_r \geq 1$.

It is proved in [Zhu05a] that for any graph G, for any real number t, if G is circular t-choosable then for $\varepsilon > 0$ and for any $(t+\varepsilon)$-circular list assignment L, G has an circular L-colouring. Conversely, if for any t-circular list assignment L, G has an circular L-colouring, then G is circular t-choosable. Therefore the circular list chromatic number of G can be defined as

$$\chi_{c,l}(G) = \inf\{t : \text{ for any } t\text{-circular list assignment } L, \; G \text{ has a circular } L\text{-colouring}\}.$$

It is easy to see that for any integer n, if G is not n-choosable, then G is not circular $(n-1)$-choosable. As a consequence, we have $\chi_{c,l}(G) \geq \chi_l(G) - 1$. However, it is surprising that the circular list chromatic number of a graph can be much large than its list chromatic number. It is proved in [Zhu05a] that for any positive integer k, for each $\varepsilon > 0$, there is a k-degenerated graph G which is not circular $(2k-\varepsilon)$-choosable. Since every k-degenerated graph is $(k+1)$-choosable, it follows that the difference $\chi_{c,l}(G) - \chi_l(G)$ can be arbitrarily large. On the other hand, we have the following result:

Theorem 12.1. *Suppose G is a finite k-degenerated graph and L is a $2k$-circular list assignment of G. Then there is a mapping f which assigns to each vertex v of G an open interval $f(v) \subseteq L(v)$ of positive length such that if $v \sim v'$, then for any $x \in f(v)$ and $x' \in f(v')$, $|x - x'|_r \geq 1$.*

As a corollary, every k-degenerated graph G has $\chi_{c,l}(G) \leq 2k$. For an arbitrary graph G, it is unknown if $\chi_{c,l}(G)$ is bounded by a multiple of $\chi_l(G)$.

Question 12.2. Is there a constant α such that for any graph G, $\chi_{c,l}(G) \leq \alpha\chi_l(G)$? If such a constant exists, what is the smallest α?

For an arbitrary graph G, it is difficult to determine $\chi_{c,l}(G)$. The question is open even for circular complete graphs.

Question 12.3. What is the circular list chromatic number of $K_{p/q}$? Is it true that $\chi_{c,l}(K_{p/q}) = p/q$?

The circular list chromatic number of odd cycles is determined in [Zhu05a]. But the question for even cycles is still open.

Theorem 12.4. *The odd cycle C_{2k+1} has $\chi_{c,l}(C_{2k+1}) = 2 + \frac{1}{k}$.*

Although a k-degenerated graph can have circular list chromatic number almost as large as $2k$, $\chi_{c,l}(G)$ is bounded by 1 plus its maximum degree.

Theorem 12.5 ([Zhu05a]). *Suppose G has maximum degree k. Then G is circular $(k+1)$-choosable.*

Circular list colouring can be used as a tool in inductive proofs in solving circular colouring problems. In an inductive proof, we may need to prove that any (p,q)-colouring of a subset X of $V(G)$ can be extended to a (p,q)-colouring of G. This would be equivalent to prove that for a certain list-size assignment ℓ, the subgraph $G - X$ is ℓ-(p,q)-choosable. For the purpose of application in such inductive proofs, one may need to consider list-size-assignments that are not constant mappings. It is a more difficult problem to characterize all colour-size-lists ℓ for which a graph G is ℓ-(p,q)-choosable. However, if G is a tree, such a characterization is given in [RZ03].

Theorem 12.6. *Suppose T is a tree, $p \geq 2q$ are positive integers and $\ell : V(T) \to \{0,1,2,\ldots,p\}$ is a colour-size-list. Then T is ℓ-(p,q)-choosable if and only if for each subtree T' of T,*

$$\sum_{v \in T'} \ell(v) \geq 2(|V(T')| - 1)q + 1.$$

The corresponding problem for cycles is also studied in [RZ03]. A sharp sufficient condition for ℓ is given in [RZ03] under which a cycle G is ℓ-$(2k+1,k)$-choosable.

Theorem 12.7. *Let $k \geq 1$ be an integer, and let $X = (x_0, x_1, \ldots, x_{n-1})$ be a cycle of length $n \geq 2k+1$. Suppose $f : V(X) \to \{0, 1, 2, \ldots, 2k+1\}$ is a colour-size-list for X. Then X is ℓ-$(2k+1,k)$-choosable if the following conditions hold:*

(1) *For each interval $[j, j']_n$ of length m, $\sum_{t \in [j,j']_n} \ell(x_t) \geq 2(m-1)k + 1$.*
(2) $\sum_{t=0}^{n-1} \ell(x_t) \geq 2nk + 1$.

Moreover, Condition (1) is necessary for X to be ℓ-$(2k+1,k)$-colourable, and in case X is an odd cycle, Condition (2) is sharp in the sense that there is a colour-size-list ℓ which satisfies (1), and $\sum_{t=0}^{n-1} \ell(x_t) = 2nk$, but X is not ℓ-$(2k+1,k)$-choosable.

13 K_4-minor Free Graphs

The class of K_4-minor free graphs have a very simple structure. Each block of a K_4-minor free graph can be obtained from C_2 by repeatedly apply series joins and parallel joins, and is called a *series-parallel graph*. The class of K_4-minor free graphs is also known as the class of partial 2-tree. Due to the simplicity of its structure, many difficult problems becomes easy when restricted to K_4-minor free graphs. One question of interest is the relation between circular chromatic number and the girth of graphs [AZ93, BKKW04, Cha01, CZ00, GGH01, HZ04, HZ00, KKŠZ05, KZ00, NZ01, NZ04, PZ02a, PZ02b, Zhou97, Zhu96a, Zhu99d]. By Corollary 10.3, in general, graphs of large girth can have arbitrary given circular chromatic number. However, if restricted to special classes of graphs, large girth graphs may be forced to have small circular chromatic number. The following result is proved in [GGH01].

Theorem 13.1. *For any integer $n \geq 4$, for any $\varepsilon > 0$, there is an integer g such that every K_n-minor free graph G of girth at least g has $\chi_c(G) \leq 2 + \varepsilon$.*

Observe that K_3-minor free graphs are forests. So Theorem 13.1 holds trivially for $n = 3$ (the requirement that $n \geq 4$ is to exclude the trivial case). Given an integer $n \geq 4$ and an $\varepsilon > 0$, let $g(n, \varepsilon)$ be the smallest integer such that every K_n-minor free graph G of girth at least g has $\chi_c(G) \leq 2 + \varepsilon$. To determine $g(n, \varepsilon)$ is very difficult problem for $n \geq 5$. For example, the best known lower and upper bound for $g(5, 1/2)$ is $9 \leq g(5, 1/2) \leq 13$. As ε gets smaller, the gap between the best known upper and lower bounds becomes larger. However, $g(4, \varepsilon)$ is completely determined. It turns out that what really matters in bounding the circular chromatic number of a K_4-minor free graph is the odd girth of G. The following results are proved in [PZ02a, PZ02b].

Theorem 13.2. *Suppose G is a K_4-minor free graph and $k \geq 1$ is an integer.*

1. *If G has odd girth at least $6k-1$ then $\chi_c(G) \leq 8k/(4k-1)$;*
2. *If G has odd girth at least $6k+1$ then $\chi_c(G) \leq (4k+1)/2k$;*
3. *If G is has odd girth at least $6k+3$ then $\chi_c(G) \leq (4k+3)/(2k+1)$.*

Theorem 13.2 strengthens a result proved in [CZ00], and the bounds given above are tight.

Theorem 13.3. *Let $k \geq 1$ be an integer, and let $\varepsilon > 0$.*

1. *There exists a series-parallel graph G of girth $6k-1$ with*
$$\chi_c(G) > 8k/(4k-1) - \varepsilon\,;$$
2. *There exists a series-parallel graph G of girth $6k+1$ with*
$$\chi_c(G) > (4k+1)/2k - \varepsilon\,;$$
3. *There exists a series-parallel graph G of girth $6k+3$ with*
$$\chi_c(G) > (4k+3)/(2k+1) - \varepsilon.$$

As a consequence of Theorems 13.2 and 13.3, we have the following

Corollary 13.4. *Suppose $1 > \varepsilon > 0$. If $1/(2k-1) > \varepsilon \geq 2/(4k-1)$, then $g(4,\varepsilon) = 6k-1$. If $1/(2k-2) > \varepsilon \geq 1/(2k-1)$, then $g(4,\varepsilon) = 6k-3$. If $2/(4k-5) > \varepsilon \geq 1/(2k-2)$, then $g(4,\varepsilon) = 6k-5$.*

As mentioned in Section 9 that for a rational r, there is a K_4-minor free graph G with $\chi_c(G) = r$ if and only if $r \in [2, 8/3] \cup \{3\}$. We say a class $\mathcal{C}$ of graphs is *universal* if for each countable partial order P, there is an injective mapping $\phi : P \to \mathcal{C}$ such that $x \leq y$ if and only if there is a homomorphism of $\phi(x)$ to $\phi(y)$. For a rational $r \in [2, 8/3] \cup \{3\}$, let $\mathcal{C}_r$ be the family of K_4-minor free graphs G with $\chi_c(G) = r$. Nešetřil and Nigussie [NN05] proved that $\mathcal{C}_r$ is universal if and only if $r \in (2, 5/2) \cup (5/2, 8/3)$.

14 Circular Flow

Suppose G is a graph and D is an orientation of G. A *flow* of G with respect to D is a mapping $f : E(D) \to \mathbb{R}$ which assigns to each arc $e = (x, y)$ of D a real number $f(e)$ such that for each cut B of G,

$$\sum_{e \in B^+} f(e) = \sum_{e \in B^-} f(e).$$

Here a cut B is the set of arcs between a subset S of G and $\bar{S} = V(G) \setminus S$, B^+ and B^- denote the sets of arcs from S to $\bar{S}$ and arcs from $\bar{S}$ to S, respectively.

For a real number $r \geq 2$, an *r-flow* of G with respect to an orientation D is a flow f such that for each arc e, $1 \leq |f(e)| \leq r-1$. If f is an r-flow of G with respect to an orientation D, and D' is the orientation obtained from D by replacing an arc $e = (x, y)$ with its opposite arc $e^{-1} = (y, x)$, then by letting $f(e^{-1}) = -f(e)$, we obtain an r-flow of G with respect to D'. So if one orientation of G has an r-flow, then every orientation has an r-flow, and we simply say G has an r-flow. The *circular flow number* $\Phi_c(G)$ of a bridgeless graph G is defined as

$$\Phi_c(G) = \min\{r : \ G \text{ has an } r\text{-flow}\}.$$

Equivalently, for integers $p \geq 2q \geq 2$, we define a (p, q)-flow to be a mapping $f : E(D) \to \{\pm q, \pm(q+1), \ldots, \pm(p-q)\}$ such that for each cut B of G,

$$\sum_{e \in B^+} f(e) = \sum_{e \in B^-} f(e).$$

Then the circular flow number of a bridgeless graph is $\Phi_c(G) = \min\{p/q : G$ has a (p, q)-flow$\}$. If $q = 1$, then a $(p, 1)$-flow f is called a *nowhere zero* p-flow. The *flow number* $\Phi(G)$ of a bridgeless graph G is defined as

$$\Phi(G) = \min\{p : \ \text{there is a nowhere zero } p\text{-flow of } G\}.$$

It follows from the definition that for any graph G,

$$\Phi(G) - 1 < \Phi_c(G) \leq \Phi(G).$$

The circular flow number of a graph can also be defined through orientations. Suppose D is an orientation of G. The *imbalance* of a cut B with respect to D is $\mathrm{Imb}_D(B) = \max\{|B|/|B^+|, |B|/|B^-|\}$. The *Cut Imbalance of* D is defined as $\mathrm{CutImb}(D) = \max\{\mathrm{Imb}_D(B) : B \text{ is a cut of } G\}$. It is proved by Goddyn, Tarsi and Zhang [GTZ98] that

$$\Phi_c(G) = \min\{\mathrm{CutImb}(D), D \text{ is an acyclic orientations of } G\}.$$

15 Coloring-flow Duality of Embedded Graphs

Coloring and flow are dual concepts in graph theory. It is proved by Tutte [Tut54] that if G is a planar graph and G^* is the geometrical dual of G, then G is k-colourable if and only if G^* admits a nowhere zero k-flow. This result can be easily extended to circular chromatic number and circular flow number: *if G is a planar graph and G^* is the geometrical dual of G, then* $\chi_c(G) = \Phi_c(G^*)$. It is natural to ask if similar results exist for graphs embedded in other surfaces.

An embedded graph G is a triple $(V(G), E(G), F(G))$, where $V(G), E(G)$ and $F(G)$ are the vertex set, edge set and face set of G, respectively. Associated with each face $R \in F(G)$, is a *boundary walk*, which is a list

$v_0, e_1, v_1, e_2, \ldots, e_k, v_k$ of vertices and edges, with $v_0 = v_k$ and with v_{i-1}, v_i be the end vertices of e_i. There are two conditions that the set of face boundaries should satisfy. First, every edge occurs precisely twice among all the face boundaries, either once in two distinct face boundaries, or twice in one. Second, for each vertex v, the edges incident with v can be enumerated $e_0, e_1, e_2, \ldots, e_{d-1}$ in such a way that for each $i \in \{0, 1, \ldots d-1\}$, e_i, e_{i+1} (summation in the index modulo d) are two consecutive edges of a face boundary. For an embedded graph G, we can construct a topological space denoted by $|G|$ as follows. If R is a face with boundary walk $v_0, e_1, v_1, e_2, \ldots, e_k, v_k$, then R is associated with a regular k-gon $\pi(R) \subseteq \mathbb{R}^2$, the vertices and edges of $\pi(R)$ correspond to those in the boundary of R. By the requirements on face boundaries, each edge of G occurs twice in the polygons $\pi(R)$ $(R \in F(G))$. The space $|G|$ is obtained from the disjoint union of these polygons by identifying both copies of every edge e. By the requirements on the face boundaries, the space $|G|$ is a *surface*, i.e., a compact 2-manifold without boundary. The *surface dual* graph G^* of G is a graph embedded in the same surface, whose vertices are the faces of G and for every $e \in E(G)$ there corresponds an edge $e^* \in E(G^*)$ connecting the two (possibly identical) faces $R, R' \in V(G^*)$ whose boundaries contain e.

The boundary of a face is a closed walk (called a *facial walk*) of G, without specific reference to the direction or the origin of the walk. By *an orientation of the faces*, we mean for each facial walk, choose one of the two directions of traversal as the positive direction of that facial walk. If the faces can be oriented in such a way that each edge e is traversed in opposite directions by the two (not necessarily distinct) facial walks incident to e, then the surface is an *orientable surface*. Otherwise the surface is *non-orientable*. The surface classification theorem states that every orientable surface is homeomorphic to S_i for some $i \geq 0$, and every non-orientable surface is homeomorphic to N_j for some $j \geq 1$, where S_i is obtained from the sphere by adding i handles, and N_j is obtained from the sphere by adding j crosscaps.

If G is embedded in an orientable surface, then an orientation of G can be transferred into an orientation of G^* as follow: Choose an orientation of the faces so that each edge e is traversed in opposite directions by the two facial walks incident to e. An edge e^* of G^* incident to R (R a vertex of G^*, but a face of G) is oriented away from R (respectively, towards R) if the direction of the corresponding edge e of G agrees (respectively, disagree) with the direction of R. For this orientation of G^*, if f is an r-tension of G, then $f'(e^*) = f(e)$ is an r-flow on G^* (where e is the arc of G corresponding to the arc e^*). Thus for a graph G embedded in an arbitrary orientable surface Σ, we still have $\Phi_c(G^*) \leq \chi_c(G)$. However, the equality $\chi_c(G) = \Phi_c(G^*)$ does not hold in general. Indeed, $\chi_c(G)$ can be arbitrarily large but $\Phi_c(G^*)$ is always bounded by 6. However, if the graph G is 'nearly' planar, then a 'relaxed duality' still exists. If a simple closed curve separates the surface Σ into two parts and one part is homeomorphic to a disc, then the curve is *contractible*. Otherwise the curve is *non-contractible*. We say a cycle of G is contractible (respectively, non-contractible) if the corresponding closed curve in Σ is contractible (respectively, non-contractible). The *edge-width* of an embedded graph G is the length of a shortest non-contractible cycle. The following result is proved in [DGMVZ05].

Theorem 15.1. *Suppose Σ is an orientable surface. For any $\varepsilon > 0$, there is an integer M such that the following holds: If G is a graph embedded in Σ with edge-width at least M and G^* is the surface dual of G, then*

$$\Phi_c(G^*) \leq \chi_c(G) \leq \Phi_c(G^*) + \varepsilon.$$

Note that if the surface is the sphere, then there is no non-contractible cycle. Hence $\Phi_c(G^*) \leq \chi_c(G) \leq \Phi_c(G^*) + \varepsilon$ holds for every $\varepsilon > 0$ and hence $\chi_c(G) = \Phi_c(G^*)$. When M is large, the condition that every non-contractible cycle has length at least M means that the graph is locally planar: to detect that the surface is not a sphere, one has to travel a long distance in the graph. So intuitively, Theorem 15.1 says that if a graph G embedded in an orientable surface is locally planar, then $\chi_c(G)$ and $\Phi_c(G^*)$ are close.

In Theorem 15.1, it is important that the surface is orientable. For example, the graph in Figure 7 can be viewed as a 5×6 grid on the Kleine bottle. The grid can be extended to a $(2k+1) \times n$ grid on the Kleine bottle, which then has edge-width $\min\{2k+1, n\}$. However, such a graph still has circular chromatic number 4, while its surface dual is an Eulerian graph and hence has circular flow number 2.

For a graph G embedded in a non-orientable surface, instead of considering flows in its surface dual G^*, it is more suitable to consider *biflows* in G^*. If G is embedded in an non-orientable surface, then no matter how the faces are oriented, there are some edges that are traversed in the same direction by both facial walks incident to it. Nevertheless, we still choose a face orientation, and then transfer an arbitrary orientation of G to an 'orientation' of G^*: An edge e^* of G^* incident to R is oriented away from R (respectively, towards R) if the orientation of corresponding edge e of G agrees with the direction of R (respectively, disagrees with the direction of R). Since the orientation of an edge e may agree (or disagree) with both facial walks incident to e, the 'orientation' defined above does not give us a directed graph. Instead, what we obtain is a *bidirected graph*, in which each edge constitutes of two half-edges (one for each end-vertex) and it may happen that the two half-edges of an edge are both directed away from (or into) its end-vertex. A *biflow* f of the bidirected graph G assigns to each edge e of G a real number $f(e)$ such that for each vertex v, $\sum_{e \in E^+(v)} f(e) = \sum_{e \in E^-(v)} f(e)$, where $E^+(v)$ (respectively, $E^-(v)$) denotes the edges directed away from (respectively, towards) v. So a biflow is the same as a flow, except that for an edge $e = xy$, we may have $e \in E^+(x)$ and $e \in E^+(y)$ (or $e \in E^-(x)$ and $e \in E^-(y)$). Bidirected graphs and biflows are introduced in [Bou83]. We define an *r-biflow* of a bidirected graph H to be a biflow f of H such that for each edge e, $1 \leq |f(e)| \leq r-1$, and define the *circular biflow number* $\Phi_c^B(H)$ of H to be the least r for which there is an r-biflow. It is conjectured that "*every bidirected graph which has a nowhere-zero biflow has a nowhere-zero* 6*-biflow*". This is equivalent to say that every bidirected graph which has a nowhere zero biflow has circular biflow number at most 6. It is proved by DeVos [DeV02] that this statement is true if 6 is replaced by 12 (and in [Bou83], the statement is proved to be true if 6 is replaced by 216). Observe that the circular biflow number is defined for bidirected graphs only. However, if G is an embedded graph, G^* is its surface dual, and the bidirected orientation on G^* is induced by an orientation of G, then different orientations of G and different choice of face orientations of G^* do not affect the circular biflow number. Indeed, the circular biflow number of G^* can be defined through the graph G. Suppose G is an oriented graph and $\phi : E(G) \to \mathbb{R}$ assigns to each edge of G a real number. A closed walk W of G is *weight balanced* if the sum of ϕ on the forward edges of W is equal to the sum of ϕ on the backward edges of W. Then a tension ϕ of G is a map $\phi : E(G) \to \mathbb{R}$ such that every closed walk is weight balanced. For an embedded graph G, a map $\phi : E(G) \to \mathbb{R}$ is a *local tension* if every facial closed walk is weight balanced. An *r-local tension* of G is a local tension ϕ of G such that for each arc e, $1 \le |\phi(e)| \le r-1$. The *local circular chromatic number* $\chi_{loc}(G)$ is defined as

$$\chi_{loc}(G) = \inf\{r : G \text{ admits a local } r\text{-tension}\}.$$

Then for any embedded graph G (either in orientable or non-orientable surface), $\chi_{loc}(G) = \Phi_c^B(G^*)$, where G^* is any bidirected graph obtained from the surface dual G^* of G induced by an orientation of G. So $\chi_{loc}(G)$ unifies both the circular flow number $\Phi_c(G^*)$ in the orientable case and the circular biflow number $\Phi_c^B(G^*)$ in the nonorientable case. With this notion, Theorem 15.1 holds for nonorientable surface as well [DGMVZ05].

Theorem 15.2. *For every surface Σ and every $\varepsilon > 0$, there exists an integer M so that every loopless Σ-embedded graph G with edge-width at least M satisfies*

$$\chi_{loc}(G) \le \chi_c(G) \le \chi_{loc}(G) + \varepsilon.$$

For an graph G embedded in an arbitrary surface, $\chi_{loc}(G)$ could be strictly smaller than $\chi_c(G)$. However, for sphere and projective plane, the two parameters coincide.

Theorem 15.3 ([DGMVZ05]). *For any graph G embedded on the sphere or projective plane we have $\chi_{loc}(G) = \chi_c(G)$.*

In 1996, Youngs discovered that every quadrangulation of the projective plane has chromatic number 2 or 4, but never 3. If we only consider the chromatic number of graphs, this is a quite isolated result. However, by considering circular chromatic number of graphs, it is shown in [DGMVZ05] that this bimodal behavior can be observed in two generic classes of embedded graphs.

Theorem 15.4. *Let G be an embedded graph.*

- *If G is even faced (i.e., each facial walk has even length) with maximum face length $2r$, then either $\chi_{loc}(G) = 2$ or $\chi_{loc}(G) \geq 2r/(r-1)$.*
- *If G is a triangulation (i.e., each facial walk has length 3), then either $\chi_{loc}(G) = 3$ or $\chi_{loc}(G) \geq 4$.*

It is known that for any surface Σ, there is an integer M such that every loopless Σ-embedded graph of edge-width at least M is 5-colourable [Tho93]; and every Σ-embedded graph of edge-width at least M and of girth at least 4 is 4-colourable. Combine these results with Theorems 15.3 and 15.4, we have the following corollary.

Corollary 15.5. *For any surface Σ and any $\varepsilon > 0$, there exists an integer M such that, for every Σ-embedded graph G with edge-width at least M,*

- *if G is even-faced, with maximum face length $2r$, then $\chi_c(G) \in [2, 2+\varepsilon] \cup [2r/(r-1), 4]$.*
- *If G is a triangulation, then $\chi_c(G) \in [3, 3+\varepsilon] \cup [4, 5]$.*

Corollary 15.5 strengthens results on the chromatic number of quadrangulations [AHNNO01, NNOŠ04], even-faced graphs [Hut95, MS02], and Eulerian triangulations [HRS02]. The upper bounds in Corollary 15.5 are sharp in the sense that there are quadrangulations G of the projective plane of arbitrarily large edge-width with $\chi_c(G) = 4$, and there are triangulations G of the projective plane of arbitrarily large edge-width with $\chi_c(G) = 5$. However, for graphs of girth at least 5, the bound can be improved. It follows from a result of Thomassen [Tho03] that if a graph G embedded in a surafce has large edge-width and girth at least 5, then G is 3-colourable. If the surface is the projective plane or the Klein bottle, then the large edge-width condition can be omitted [TW04, Tho94].

By Tutte's 5-Flow Conjecture, if a loopless graph G is embedded in an orientable surface with cogirth at least 2, then $\chi_{loc}(G) \leq 5$. By Bouchet's 6-Flow Conjecture, if a loopless graph G is embedded in any surface with cogirth at least 2, then $\chi_{loc}(G) \leq 6$. If the embedded graph G has large edge-width, then it is likely that there are better upper bounds on $\chi_{loc}(G)$. A detailed discussion about such conjectures can be found in [DGMVZ05]. For a given surface Σ, let $\chi_w(\Sigma) = \lim_{w \to \infty} \sup\{\chi_c(G) : G \text{ is embedded in } \Sigma \text{ with edge-width at least } w\}$. It follows from the above mentioned results that if Σ is nonorientable, then $\chi_w(\Sigma) = 5$. It is conjectured by Goddyn [God05] that if Σ is an orientable surface, then $\chi_w(\Sigma) = 4$. Equivalently, the conjecture says that for any orientable surface Σ, for any $\varepsilon > 0$ there is an integer M such that every Σ-embedded graph with edge-width at least M has $\chi_c(G) < 4 + \varepsilon$. By using Theorem 15.3, we can see that this conjecture is weaker than a conjecture of Grünbaum [Grü69], which says that if Σ is an orientable surface, then every Σ-embedded graph of edge-width at least 3 has $\chi_{loc}(G) \leq 4$.

16 Circular Flow Number

The fundamental problems concerning flow number and circular flow number are the possible values of the flow number and the circular flow number of graphs v.s. edge connectivity. If G has an edge cut, then G has no r-flow for any r, and $\Phi(G)$, $\Phi_c(G)$ are not defined (or defined to be ∞). There are three famous conjectures of Tutte [Tut56, Tut66, Ste76] relating the flow number of a graph to its edge connectivity.

5-Flow Conjecture. *Every bridgeless graph G has $\Phi(G) \leq 5$.*

4-Flow Conjecture. *Every Petersen minor free bridgeless graph G has*

$$\Phi(G) \leq 4.$$

3-Flow Conjecture. *Every 4-edge connected graph G has $\Phi(G) \leq 3$.*

All the three conjectures remain open. On the other hand, the following is a combination of results in [Mos97, PZ03, Zhu99b].

Theorem 16.1. *For any rational $r \in [2,5]$, there is a graph G with $\Phi_c(G) = r$. For any $r \in [2,4]$, there is a planar graph (and hence a Petersen minor free graph) G with $\Phi_c(G) = r$. For any $r \in [2,3]$, there is a 4-edge connected (planar) graph G with $\Phi_c(G) = r$.*

Another important conjecture, which generalizes both the 5-flow conjecture and 3-flow conjectrue, and which relates connectivity and circular flow number of a graph, is proposed by Jaeger [Jae88]:

$(2 + 1/k)$-Flow Conjecture. *Every $4k$-edge connected graph G has*

$$\Phi_c(G) \leq 2 + 1/k.$$

The conjecture says that graphs without small edge cut have small circular flow number. It seems that to ensure a graph to have small circular flow number, what really matters is that there are no small odd edge cut. Define the *odd edge connectivity* of a graph G to be the smallest odd number k for which there is an edge cut of cardinality k. Based on this observation, Zhang [Zha02] modified Jaeger's conjecture to the following stronger version:

$(2 + 1/k)$-Flow Conjecture (strong version). *Every graph G with odd edge connectivity at least $4k + 1$ has $\Phi_c(G) \leq 2 + 1/k$.*

The $k = 1$ case of the conjecture is exactly the 3-flow conjecture. The $k = 2$ case implies the 5-Flow Conjecture. To see this, we assume the $(2+1/k)$-Flow Conjecture is true for $k = 2$. To prove that the 5-Flow Conjecture is true, it suffices to show that any 3-edge connected cubic graph G has a nowhere zero 5-flow [Zha97]. Replace each edge of G by three parallel edges, the resulting graph G' is 9-edge connected, and hence has $\Phi_c(G') \leq 5/2$. Let D be an orientation of G and let D' be the orientation of G' obtained from D by replacing each arc of D with three arcs of the same direction. Let f be a $(5,2)$-flow on D'. For each arc a of D, let a_1, a_2, a_3 be the three parallel arcs in D' that replace a. Then $g(a) = [f(a_1) + f(a_2) + f(a_3)]_5$ is a nowhere zero Z_5-flow on D, which implies that $\Phi(G) \leq 5$ [Zha97].

Although the 5-flow, 4-flow and 3-Flow Conjectures are open, their restrictions to planar graphs have been proved. For the $(2+1/k)$-Flow Conjecture, its restriction to planar graphs also remains open. For planar graphs, the circular flow number of G is equal to the circular chromatic number of its geometric dual graph G^*. The odd edge connectivity of G is equal to the odd girth (i.e., the length of a shortest odd cycle) of G^*. So the restriction of the strong version of the $(2+1/k)$-Flow Conjecture to planar graphs is equivalent to the following:

$(2+1/k)$-Flow Conjecture for planar graphs (strong version). *Every planar graph G with odd girth at least $4k+1$ has $\chi_c(G) \leq 2 + 1/k$.*

The circular chromatic number of planar graphs of large girth or large odd girth has been studied in [BKKW04, GGH01, KZ00, FJMŠ01, Zhu01a]. The currently best known result is the following theorem proved in [BKKW04]:

Theorem 16.2. *If G is a planar graph of odd girth at least $\frac{20k-2}{3}$, then $\chi_c(G) \leq 2 + 1/k$.*

The following weaker version of $(2+1/k)$-Flow Conjecture is proposed by Seymour (cf. [Zha02]):

$(2+1/k)$-Flow Conjecture (weak version). *For any $\varepsilon > 0$, there is an integer $n = n(\varepsilon)$ such that every graph G of girth (or odd girth) at least n has $\Phi_c(G) \leq 2 + \varepsilon$.*

This conjecture also remains open. Theorem 16.2 implies that the weak version of $(2+1/k)$-Flow Conjecture holds for planar graphs. Zhang [Zha02] proved that for any given surface S, the weak version of $(2+1/k)$-Flow Conjecture holds for graphs embedded in S.

Theorem 16.3. *Let S be any given surface and ε be a positive real number. There is an integer $n = n_S(\varepsilon)$ such that any graph G with odd edge connectivity at least n has $\Phi_c(G) \leq 2 + \varepsilon$.*

Circular flow number of random graphs is discussed in [Sud01]. It turns out that $(2+1/k)$-Flow Conjecture is true almost surely for random graphs.

Theorem 16.4. *In the random graph process which adds a uniformly chosen edge at each step, almost surely the graph has $\Phi_c(G) \leq 2 + 1/k$ as soon as the minimum degree of the graph is at least $2k$.*

Let $G(n,p)$ denote a random graph with n vertices in which each pair of vertices is joined by an edge with probability p.

Theorem 16.5. *Let $k \geq 1$ be an integer and let $\omega(n)$ be any function tending to infinity with n. Then*

- *if $p = (\ln n + (2k-1)\ln\ln n + \omega(n))/n$, then almost surely $\Phi_c(G(n,p)) \leq 2 + 1/k$;*
- *if $p = (\ln n + (2k-1)\ln\ln n - \omega(n))/n$, then almost surely $\Phi_c(G(n,p)) > 2 + 1/k$.*

Random regular graphs are also considered in [Sud01]. When $d = 3$, it is known that an element of $G_{n,d}$ almost surely contains an odd cycle, and therefore has no nowhere zero 3-flow. It is shown in [Sud01] that for odd $d \geq 11$, $G_{n,d}$ almost surely has a nowhere zero 3-flow. The corresponding result for $d = 5$ is posed as an open problem, for which Tutte's conjecture implies a positive answer.

It is also a long standing open question as whether there is an integer n such that every graph G of (odd) edge connectivity at least n has $\Phi(G) \leq 3$. It is known [Zha97] 4-edge connected graphs has $\Phi(G) \leq 4$. For 6-edge connected graphs, there is a slightly better result [GG02].

Theorem 16.6. *If G is 6-edge connected, then $\Phi_c(G) < 4$.*

Possible values of the circular flow number of regular graphs are studied by Steffen [Ste01]. The following result is proved:

Theorem 16.7. *If G is a $(2k+1)$-regular graphs, then either $\Phi_c(G) \geq 2+\frac{2}{2k-1}$ or G is bipartite and in which case $\Phi_c(G) = 2 + \frac{1}{k}$.*

References

[AZ93] H. L. Abbott and B. Zhou, The star chromatic number of a graph, *J. Graph Theory*, 17(3):349–360, 1993.

[AGGHTZ05] P. Afshani, M. Ghandehari, M. Ghandehari, H. Hatami, R. Tusserkani, and X. Zhu, Circular chromatic index of graphs of maximum degree 3, *J. Graph Theory*, 49:325–335, 2005.

[AHT03] P. Afshani, H. Hatami, and R. Tusserkani, A note on circular chromatic number of Mycielski graphs, Manuscript, 2003.

[AHNNO01] D. Archdeacon, J. Hutchinson, A. Nakamoto, S. Negam, and K. Ota, Chromatic numbers of quadrangulations on closed surfaces, *J. Graph Theory*, 37(2):100–114, 2001.

[BK03] E. Babson and D. N. Kozlov, Complexes of graph homomorphisms, Preprint, arXiv:math.CO/0310056, 2003.

[BH02] J. Bang-Jensen and J. Huang, Convex-round graphs are circular-perfect, *J. Graph Theory*, 40(3):182–194, 2002.

[Bar86] V. C. Barbosa, *Concurreny systems with neighbourhood constraints*, Ph.D. Dissertation, Computer Science Dept., University of Carlifornia, Los Angeles, 1986.

[BG87] V. C. Barbosa and E. Gafni, Concurreny in heavily loaded neighbourhood-constrained systems (condensed version), In *Proceedings of the 7th International Conference on Distributed Computing Systems (Washington, D.C., Sept. 1987)*. IEEE Computer Society Press, New York, 1987.

[BG89] V. C. Barbosa and E. Gafni, Concurreny in heavily loaded neighbourhood-constrained systems, *ACM Trans. on Programming Languages and Systems*, 11:562–584, 1989.

[Bar02] V. C. Barbosa, The interleaved multichromatic number of a graph, *Ann. Comb.*, 6(3-4):249–256, 2002.

[BZ98] B. L. Bauslaugh and X. Zhu, Circular colourings of infinite graphs, *Bull. Inst. Combin. Appl.*, 24:79–80, 1998.

[BFJKM04] D. Bokal, G. Fijavž, M. Juvan, P. Mark Kayll, and B. Mohar, The circular chromatic number of a digraph, *J. Graph Theory*, 46(3):227–240, 2004.

[BH90] A. J. Bondy and P. Hell, A note on the star chromatic number, *J. Graph Theory*, 14(4):479–482, 1990.

[BKKW04] O. V. Borodin, S.-J. Kim, A. V. Kostochka, and D. B. West, Homomorphisms from sparse graphs with large girth, *J. Combin. Theory Ser. B*, 90(1):147–159, 2004. Dedicated to Adrian Bondy and U. S. R. Murty.

[Bou83] A. Bouchet, Nowhere-zero integral flows on a bidirected graph, *J. Combin. Theory Ser. B*, 34(3):279–292, 1983.

[BH02] R. C. Brewster and P. Hell, Homomorphisms to powers of digraphs, *Discrete Math.*, 244(1-3):31–41, 2002. Algebraic and topological methods in graph theory (Lake Bled, 1999).

[BEL76] S. A. Burr, P. Erdős, and L. Lovász, On graphs of Ramsey type, *Ars Combin.*, 1(1):167–190, 1976.

[Cha01] G. J. Chang, A survey on circular chromatic numbers of graphs, In *First International Congress of Chinese Mathematicians (Beijing, 1998)*, volume 20 of *AMS/IP Stud. Adv. Math.*, pages 497–502. Amer. Math. Soc., Providence, RI, 2001.

[CHZ98] G. J. Chang, L. Huang, and X. Zhu, Circular chromatic numbers and fractional chromatic numbers of distance graphs, *European J. Combin.*, 19(4):423–431, 1998.

[CHZ99] G. J. Chang, L. Huang, and X. Zhu, Circular chromatic numbers of Mycielski's graphs, *Discrete Math.*, 205(1-3):23–37, 1999.

[CLZ99] G. J. Chang, D. D.-F. Liu, and X. Zhu, Distance graphs and T-coloring, *J. Combin. Theory Ser. B*, 75(2):259–269, 1999.

[CZ00] C. Chien and X. Zhu, The circular chromatic number of series-parallel graphs with large girth, *J. Graph Theory*, 33(4):185–198, 2000.

[CPW05] S. Coulonges, A. Pêcher, and A. K. Wagler, Triangle-free strongly circular perfect graphs, Manuscript, 2005.

[CPW05a] S. Coulonges, A. Pêcher, and A. K. Wagler, On strongly circular perfectness, Manuscript, 2005.

[DZ96] W. A. Deuber and X. Zhu, Circular colorings of weighted graphs, *J. Graph Theory*, 23(4):365–376, 1996.

[DeV02] M. DeVos, Flows on bidirected graphs, Manuscript, 2002.

[DGMVZ05] M. DeVos, L. Goddyn, B. Mohar, D. Vertigan, and X. Zhu, Coloring-flow duality of embedded graphs, *Trans. Amer. Math. Soc.*, 357(10):3993–4016 (electronic), 2005.

[DSW85] D. Duffus, B. Sands, and R. E. Woodrow, On the chromatic number of the product of graphs, *J. Graph Theory*, 9(4):487–495, 1985.

[ES85] M. H. El-Zahar and N. Sauer, The chromatic number of the product of two 4-chromatic graphs is 4, *Combinatorica*, 5(2):121–126, 1985.

[EHK98] T. Emden-Weinert, S. Hougardy, and B. Kreuter, Uniquely colourable graphs and the hardness of colouring graphs of large girth, *Combin. Probab. Comput.*, 7(4):375–386, 1998.

[Fan04] G. Fan, Circular chromatic number and Mycielski graphs, *Combinatorica*, 24(1):127–135, 2004.

[FHM03] T. Feder, P. Hell, and B. Mohar, Acyclic homomorphisms and circular colorings of digraphs, *SIAM J. Discrete Math.*, 17(1):161–169 (electronic), 2003.

[FJMŠ01] G. Fijavž, M. Juvan, B. Mohar, and R. Škrekovski, Circular colorings of planar graphs with prescribed girth, preprint, 2001.

[GG02] A. Galluccio and L. A. Goddyn, The circular flow number of a 6-edge connected graph is less than four, *Combinatorica*, 22(3):455–459, 2002.

[GGH01] A. Galluccio, L. A. Goddyn, and P. Hell, High-girth graphs avoiding a minor are nearly bipartite, *J. Combin. Theory Ser. B*, 83(1):1–14, 2001.

[GZ96] G. Gao and X. Zhu, Star-extremal graphs and the lexicographic product, *Discrete Math.*, 152(1-3):147–156, 1996.

[GKNT06] M. Ghebleh, D. Král', S. Norine, and R. Thomas, Circular chromatic index of flower snarks, *J. Graph Theory*, to appear.

[God05] L. Goddyn, Wide embedded graphs behave chromatically like plane or projective plane, In *Graph Theory*, Mathematisches Forschungsinstitut Oberwolfach, Report No. 3. 2005.

[GTZ98] L. A. Goddyn, M. Tarsi, and C.-Q. Zhang, On (k, d)-colorings and fractional nowhere-zero flows, *J. Graph Theory*, 28(3):155–161, 1998.

[Grü69] B. Grünbaum, Conjecture 6, in "Recent progress in combinatorics", *Proceedings of the Third Waterloo Conference on Combinatorics*, May 1968. Edited by W. T. Tutte. Academic Press, New York, 1969.

[Gui93] D. R. Guichard, Acyclic graph coloring and the complexity of the star chromatic number, *J. Graph Theory*, 17(2):129–134, 1993.

[GJS04] A. Gyárfás, T. Jensen, and M. Stiebitz, On graphs with strongly independent color-classes, *J. Graph Theory*, 46(1):1–14, 2004.

[HK04] A. Hackmann and A. Kemnitz, The circular chromatic index, *Discrete Math.*, 286(1-2):89–93, 2004.

[HHMN88] R. Häggkvist, P. Hell, D. J. Miller, and V. Neumann Lara, On multiplicative graphs and the product conjecture, *Combinatorica*, 8(1):63–74, 1988.

[HT97] G. Hahn and C. Tardif, Graph homomorphisms: structure and symmetry, In *Graph symmetry (Montreal, PQ, 1996)*, volume 497 of *NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci.*, pages 107–166. Kluwer Acad. Publ., Dordrecht, 1997.

[HZ03b] H. Hajiabolhassan and X. Zhu, Circular chromatic number and Mycielski construction, *J. Graph Theory*, 44(2):106–115, 2003.

[HZ03c] H. Hajiabolhassan and X. Zhu, Circular chromatic number of Kneser graphs, *J. Combin. Theory Ser. B*, 88(2):299–303, 2003.

[HZ03a] H. Hajiabolhassan and X. Zhu, Circular chromatic number of subgraphs, *J. Graph Theory*, 44(2):95–105, 2003.

[HZ04] H. Hajiabolhassan and X. Zhu, Sparse H-colourable graphs of bounded maximum degree, *Graphs Combin.*, 20(1):65–71, 2004.

[HT04] H. Hatami and R. Tusserkani, On the complexity of the circular chromatic number, *J. Graph Theory*, 47(3):226–230, 2004.

[Hat05] H. Hatami, Random cubic graphs are not homomorphic to the cycle of size 7, *J. Combin. Theory Ser. B*, 93(2):319–325, 2005.

[JZ05] H. Hatami and X. Zhu, The fractional chromatic number of graphs of maximum degree at most three, Manuscript, 2005.

[Hed66] S. Hedetniemi, Homomorphisms and graph automata, University of Michigan Technical Report 03105-44-T, 1966.

[HN89] P. Hell and J. Nešetřil, On the complexity of H-coloring, *J. Combin. Theory Ser. B*, 48(1):92–110, 1990.

[HN04] P. Hell and J. Nešetřil, *Graphs and Homomorphisms*, Clarendon Press · Oxford, Oxford, 2004.

[HZ00] P. Hell and X. Zhu, The circular chromatic number of series-parallel graphs, *J. Graph Theory*, 33(1):14–24, 2000.

[Hof60] A. J. Hoffman, Some recent applications of the theory of linear inequalities to exremal combinatorial analysis, *Combinatorial Analysis: Proceedings of the Tenth Symposium in Applied Mathematics of the American Mathematical Society. R. Bellman and M. Hall Jr., Eds.*, pages 113–128, 1960.

[HC00] L. Huang and G. J. Chang, Circular chromatic numbers of distance graphs with distance sets missing multiples, *European J. Combin.*, 21(2):241–248, 2000.

[HRS02] J. Hutchinson, R. Bruce Richter, and P. Seymour, Colouring Eulerian triangulations, *J. Combin. Theory Ser. B*, 84(2):225–239, 2002.

[Hut95] J. P. Hutchinson, Three-coloring graphs embedded on surfaces with all faces even-sided, *J. Combin. Theory Ser. B*, 65(1):139–155, 1995.

[JW80] F. Jaeger and T. Swart, Conjecture 1 in "Combinatorics 79" (M. Deza and I. G. Rosenberg, Eds.), *Ann. Discrete Math.*, 9:305, 1980.

[Jae88] F. Jaeger, Nowhere-zero flow problems, In *Selected topics in graph theory, 3*, pages 71–95. Academic Press, San Diego, CA, 1988.

[Jhan05] W. Jhan, *Circular chromatic indexes of generalized necklaces*, Master Thesis, National Sat-yen Sun University, 2005.

[JHS97] A. Johnson, F. C. Holroyd, and S. Stahl, Multichromatic numbers, star chromatic numbers and Kneser graphs, *J. Graph Theory*, 26(3):137–145, 1997.

[KKŠ05] T. Kaiser, D. Král', and R. Škrekovski, A revival of the girth conjecture, *J. Combin. Theory Ser. B*, 92(1):41–53, 2004.

[KKŠZ05] T. Kaiser, D. Král', R. Škrekovski, and X. Zhu, The circular chromatic index of graphs of large girth, *J. Combin. Theory Ser. B*, submitted.

[KZ00] W. Klostermeyer and C. Q. Zhang, $(2+\varepsilon)$-coloring of planar graphs with large odd-girth, *J. Graph Theory*, 33(2):109–119, 2000.

[Kne55] M. Kneser, Aufgabe 360, *Jahresbericht der Deutschen Mathematiker-Vereinigung, 58:2. Abteilung, S. 27*, 1955.

[Koc96] M. Kochol, Snarks without small cycles, *J. Combin. Theory Ser. B*, 67(1):34–47, 1996.

[LLGS03] P. Che Bor Lam, W. Lin, G. Gu, and Z. Song, Circular chromatic number and a generalization of the construction of Mycielski, *J. Combin. Theory Ser. B*, 89(2):195–205, 2003.

[LT00] B. Larose and C. Tardif, Hedetniemi's conjecture and the retracts of a product of graphs, *Combinatorica*, 20(4):531–544, 2000.

[LT01] B. Larose and C. Tardif, Strongly rigid graphs and projectivity, *Mult.-Valued Log.*, 7(5-6):339–361, 2001. Ivo G. Rosenberg's 65th birthday, Part 2.

[LT02] B. Larose and C. Tardif, Projectivity and independent sets in powers of graphs, *J. Graph Theory*, 40(3):162–171, 2002.

[LPZ03] S.-C. Liaw, Z. Pan, and X. Zhu, Construction of K_n-minor free graphs with given circular chromatic number, *Discrete Math.*, 263(1-3):191–206, 2003.

[LL02] K.-W. Lih and D. D.-F. Liu, Circular chromatic numbers of some reduced Kneser graphs, *J. Graph Theory*, 41(1):62–68, 2002.

[LLZ99] K.-W. Lih, D. D.-F. Liu, and X. Zhu, Star extremal circulant graphs, *SIAM J. Discrete Math.*, 12(4):491–499 (electronic), 1999.

[Lin03] W. Lin, Some star extremal circulant graphs, *Discrete Math.*, 271(1-3):169–177, 2003.

[LG04] W. Lin and G. Gu, A necessary and sufficient condition for a vertex-transitive graph to be star extremal, *J. Southeast Univ. (English Ed.)*, 20(3):374–377, 2004.

[LLS05] W. Lin, P. Che Bor Lam, and Z. Song, Circular chromatic numbers of some distance graphs, *Discrete Math.*, 292(1-3):119–130, 2005.

[LZ99] D. D.-F. Liu and X. Zhu, Distance graphs with missing multiples in the distance sets, *J. Graph Theory*, 30(4):245–259, 1999.

[Liu04] D. D.-F. Liu, Circular chromatic number for iterated Mycielski graphs, *Discrete Math.*, 285(1-3):335–340, 2004.

[LZ05] D. D.-F. Liu and X. Zhu, Coloring the cartesian sum of graphs, Manuscript, 2005.

[Lov78] L. Lovász, Kneser's conjecture, chromatic number, and homotopy, *J. Combin. Theory Ser. A*, 25(3):319–324, 1978.

[ŁN04] T. Łuczak and J. Nešetřil, Note on projective graphs, *J. Graph Theory*, 47(2):81–86, 2004.

[Meu05] F. Meunier, A topological lower bound for the circualr chromatic number of schrijver graphs, *J. Graph Theory*, 49:257–261, 2005.

[Moh03] B. Mohar, Circular colorings of edge-weighted graphs, *J. Graph Theory*, 43(2):107–116, 2003.

[Moh05] B. Mohar, Hajós theorem for colorings of edge-weighted graphs, *Combinatorica*, 25(1):65–76, 2005.

[MS02] B. Mohar and P. D. Seymour, Coloring locally bipartite graphs on surfaces, *J. Combin. Theory Ser. B*, 84(2):301–310, 2002.

[Mos95] D. Moser, The star chromatic number of line graphs, In *Graph theory, combinatorics, and algorithms, Vol. 1, 2 (Kalamazoo, MI, 1992)*, Wiley-Intersci. Publ., pages 819–823. Wiley, New York, 1995.

[Mos97] D. Moser, The star-chromatic number of planar graphs, *J. Graph Theory*, 24(1):33–43, 1997.

[Mül79] V. Müller, On colorings of graphs without short cycles, *Discrete Math.*, 26(2):165–176, 1979.

[Nad06] A. Nadolski, The circular chromatic index of some class 2 graphs, *Discrete Math.*, to appear.

[NNOŠ04] A. Nakamoto, S. Negami, K. Ota, and J. Širáň, Planar triangulations which quadrangulate other surfaces, *European J. Combin.*, 25(6):817–833, 2004.

[NN05] J. Nešetřil and Y. Nigussie, Density of universal classes in $\mathcal{G}/K_4$, Manuscript, 2005.

[Neš99a] J. Nešetřil, Aspects of structural combinatorics (graph homomorphisms and their use), *Taiwanese J. Math.*, 3(4):381–423, 1999.

[Neš99b] J. Nešetřil, The homomorphism structure of classes of graphs, *Combin. Probab. Comput.*, 8(1-2):177–184, 1999. Recent trends in combinatorics (Mátraháza, 1995).

[NZ01] J. Nešetřil and X. Zhu, Construction of sparse graphs with prescribed circular colorings, *Discrete Math.*, 233(1-3):277–291, 2001. Graph theory (Prague, 1998).

[NZ04] J. Nešetřil and X. Zhu, On sparse graphs with given colorings and homomorphisms, *J. Combin. Theory Ser. B*, 90(1):161–172, 2004. Dedicated to Adrian Bondy and U. S. R. Murty.

[PZ02a] Z. Pan and X. Zhu, The circular chromatic number of series-parallel graphs of large odd girth, *Discrete Math.*, 245(1-3):235–246, 2002.

[PZ02b] Z. Pan and X. Zhu, Tight relation between the circular chromatic number and the girth of series-parallel graphs, *Discrete Math.*, 254(1-3):393–404, 2002.

[PZ03] Z. Pan and X. Zhu, Construction of graphs with given circular flow numbers, *J. Graph Theory*, 43(4):304–318, 2003.

[PZ04] Z. Pan and X. Zhu, Density of the circular chromatic numbers of series-parallel graphs, *J. Graph Theory*, 46(1):57–68, 2004.

[PZ05] Z. Pan and X. Zhu, Graphs of large girthwith prescribed partial circular colourings, *Graphs Combin.*, 21:119–129, 2005.

[PZ06] Z. Pan and X. Zhu, On minimal circular imperfect graphs, In preparation, 2005.

[PZ05] A. Pêcher and X. Zhu, On the circular chromatic number of circular partitionable graphs, *J. Graph Theory*, to appear.

[PWZ05] A. Pêcher, A.K. Wagler, and X. Zhu, Three classes of minimal circular-imperfect graphs, Manuscript, 2005.

[PR81] S. Poljak and V. Rödl, On the arc-chromatic number of a digraph, *J. Combin. Theory Ser. B*, 31(2):190–198, 1981.

[Pol92] S. Poljak, Coloring digraphs by iterated antichains, *Comment. Math. Univ. Carolin.*, 32(2):209–212, 1991.

[RZ03] A. Raspaud and X. Zhu, List circular coloring of trees and cycles, Manuscript, 2003.

[SZ92] N. Sauer and X. Zhu, An approach to Hedetniemi's conjecture, *J. Graph Theory*, 16(5):423–436, 1992.

[Sau01] N. Sauer, Hedetniemi's conjecture—a survey, *Discrete Math.*, 229(1-3):261–292, 2001. Combinatorics, graph theory, algorithms and applications.

[Sch78] A. Schrijver, Vertex-critical subgraphs of Kneser graphs, *Nieuw Arch. Wisk. (3)*, 26(3):454–461, 1978.

[She95] F. B. Shepherd, Applying Lehman's theorem to packing problems, *Math. Programming*, 71:353–367, 1995.

[ST04] G. Simonyi and G. Tardos, Local chromatic number, ky fan's theorem, and circualr colorings, Preprint, arXiv:math.CO/0407075, 2004.

[Soh05] C.-A. Soh, The circular chromatic number of planar digraphs, Manuscript, 2004.

[Ste01] E. Steffen, Circular flow numbers of regular multigraphs, *J. Graph Theory*, 36(1):24–34, 2001.

[SZ96] E. Steffen and X. Zhu, Star chromatic numbers of graphs, *Combinatorica*, 16(3):439–448, 1996.

[Ste76] R. Steinberg, Grotzsch's theorem dualized, *M. Math Thesis, University of Waterloo, Ontario, Canada*, 1976.

[Sti85] M. Stiebitz, *Beiträge zur theorie der färbungskritischen graphen*, Habilitation, TH Illmenau, 1985.

[Sud01] B. Sudakov, Nowhere-zero flows in random graphs, *J. Combin. Theory Ser. B*, 81(2):209–223, 2001.

[Tar01b] C. Tardif, The chromatic number of the product of two graphs is at least half the minimum of the fractional chromatic numbers of the factors, *Comment. Math. Univ. Carolin.*, 42(2):353–355, 2001.

[Tar04] C. Tardif, Chromatic numbers of products of tournaments: fractional aspects of Hedetniemi's conjecture, In *Graphs, morphisms and statistical physics*, volume 63 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 171–175. Amer. Math. Soc., Providence, RI, 2004.

[Tar05b] C. Tardif, The fractional chromatic number of the catoegorial product of graphs, Preprint, 2005.

[Tar05a] C. Tardif, Multiplicative graphs and semi-lattice endmorphisms in the catoegory of graphs, *J. Combin. Theory Ser. B*, to appear.

[TZ02b] C. Tardif and X. Zhu, On Hedetniemi's conjecture and the colour template scheme, *Discrete Math.*, 253(1-3):77–85, 2002. Combinatorics and algorithms (Hsin Chu/Kaohsiung, 2000).

[TW04] R. Thomas and B. Walls, Three-coloring Klein bottle graphs of girth five, *J. Combin. Theory Ser. B*, 92(1):115–135, 2004.

[Tho93] C. Thomassen, Five-coloring maps on surfaces, *J. Combin. Theory Ser. B*, 59(1):89–105, 1993.

[Tho94] C. Thomassen, Grötzsch's 3-color theorem and its counterparts for the torus and the projective plane, *J. Combin. Theory Ser. B*, 62(2):268–279, 1994.

[Tho03] C. Thomassen, The chromatic number of a graph of girth 5 on a fixed surface, *J. Combin. Theory Ser. B*, 87(1):38–71, 2003. Dedicated to Crispin St. J. A. Nash-Williams.

[Tur83] D. Turzík, A note on chromatic number of direct product of graphs, *Comment. Math. Univ. Carolin.*, 24(3):461–463, 1983.

[Tut54] W. T. Tutte, A contribution to the theory of chromatic polynomials, *Canadian J. Math.*, 6:80–91, 1954.

[Tut56] W. T. Tutte, A class of Abelian groups, *Canad. J. Math.*, 8:13–28, 1956.

[Tut66] W. T. Tutte, On the algebraic theory of graph colorings, *J. Combinatorial Theory*, 1:15–50, 1966.

[Vin88] A. Vince, Star chromatic number, *J. Graph Theory*, 12(4):551–559, 1988.

[Wag03] A.K. Wagler, Antiwebs are rank-perfect, *Quarterly Journal of the Belgian, French and Italian OR Societies*, 2003.

[Wel84] E. Welzl, Symmetric graphs and interpretations, *J. Combin. Theory Ser. B*, 37(3):235–244, 1984.

[WZ06] D. West and X. Zhu, Circular chromatic indexes of cartesian product of graphs, Manuscript, 2004.

[WL04] J. Wu and W. Lin, Circular chromatic numbers and fractional chromatic numbers of distance graphs with distance sets missing an interval, *Ars Combin.*, 70:161–168, 2004.

[Xu05c] B. Xu, An analogue of dirac's theorem on circular critical graphs, Manuscript, 2005.

[Xu05a] B. Xu, A family of minimally circular-imperfect series-parallel graphs, Manuscript, 2005.

[Xu05b] B. Xu, On minimally circular-imperfect graphs with a major vertex, Manuscript, 2005.

[Yan05] C. Yang, *Perfectness of the complements of circular complete graphs*, Master Thesis, National Sat-yen Sun University, 2005.

[Yeh05] H.-G. Yeh, A dynamic view of circular colorings, Manuscript, 2005.

[YZ05] H.-G. Yeh and X. Zhu, Resource-sharing system scheduling and circular chromatic number, *Theoret. Comput. Sci.*, 332(1-3):447–460, 2005.

[Zha97] C.-Q. Zhang, *Integer flows and cycle covers of graphs*, volume 205 of *Monographs and Textbooks in Pure and Applied Mathematics*, Marcel Dekker Inc., New York, 1997.

[Zha02] C.-Q. Zhang, Circular flows of nearly Eulerian graphs and vertex-splitting, *J. Graph Theory*, 40(3):147–161, 2002.

[Zhou97] B. Zhou, Some theorems concerning the star chromatic number of a graph, *J. Combin. Theory Ser. B*, 70(2):245–258, 1997.

[Zhu92] X. Zhu, Star chromatic numbers and products of graphs, *J. Graph Theory*, 16(6):557–569, 1992.

[Zhu96b] X. Zhu, On the bounds for the ultimate independence ratio of a graph, *Discrete Math.*, 156(1-3):229–236, 1996.

[Zhu96a] X. Zhu, Uniquely H-colorable graphs with large girth, *J. Graph Theory*, 23(1):33–41, 1996.

[Zhu98] X. Zhu, A survey on Hedetniemi's conjecture, *Taiwanese J. Math.*, 2(1):1–24, 1998.

[Zhu99c] X. Zhu, Circular colouring and graph homomorphism, *Bull. Austral. Math. Soc.*, 59(1):83–97, 1999.

[Zhu99d] X. Zhu, Construction of uniquely H-colorable graphs, *J. Graph Theory*, 30(1):1–6, 1999.

[Zhu99a] X. Zhu, Graphs whose circular chromatic number equals the chromatic number, *Combinatorica*, 19(1):139–149, 1999.

[Zhu99b] X. Zhu, Planar graphs with circular chromatic numbers between 3 and 4, *J. Combin. Theory Ser. B*, 76(2):170–200, 1999.

[Zhu99e] X. Zhu, A simple proof of Moser's theorem, *J. Graph Theory*, 30(1):19–26, 1999.

[Zhu00] X. Zhu, Circular chromatic number and graph minors, *Taiwanese J. Math.*, 4(4):643–660, 2000.

[Zhu01b] X. Zhu, An analogue of Hajós' theorem for the circular chromatic number, *Proc. Amer. Math. Soc.*, 129(10):2845–2852 (electronic), 2001.

[Zhu01c] X. Zhu, Circular chromatic number: a survey, *Discrete Math.*, 229(1-3):371–410, 2001. Combinatorics, graph theory, algorithms and applications.

[Zhu01a] X. Zhu, Circular chromatic number of planar graphs of large odd girth, *Electron. J. Combin.*, 8(1):Research Paper 25, 11 pp. (electronic), 2001.

[Zhu02a] X. Zhu, Circular chromatic number of distance graphs with distance sets of cardinality 3, *J. Graph Theory*, 41(3):195–207, 2002.

[Zhu02b] X. Zhu, Circular colouring and orientation of graphs, *J. Combin. Theory Ser. B*, 86(1):109–113, 2002.

[Zhu02c] X. Zhu, The fractional chromatic number of the direct product of graphs, *Glasg. Math. J.*, 44(1):103–115, 2002.

[Zhu03a] X. Zhu, An analogue of Hajós' theorem for the circular chromatic number, II, *Graphs Combin.*, 19(3):419–432, 2003.

[Zhu04b] X. Zhu, The circular chromatic number of induced subgraphs, *J. Combin. Theory Ser. B*, 92(1):177–181, 2004.

[Zhu04] X. Zhu, Perfect graphs for generalized colouring - circular perfect graphs, *AMS, DIMACS series: Proceedings of the DIMACS/DIMATIA Workshop on Graphs, Morphisms and Statistical Physics*, 63:177–193, 2004.

[Zhu05a] X. Zhu, Circular choosability of graphs, *J. Graph Theory*, 48(3):210–218, 2005.

[Zhu05b] X. Zhu, Circular perfect graphs, *J. Graph Theory*, 48(3):186–209, 2005.

Part VI

Graph Embeddings

Regular Embeddings of Multigraphs

Hubert de Fraysseix and Patrice Ossona de Mendez

Centre d'Analyse et de Mathématiques Sociales, CNRS, UMR 8557, 54 Bd Raspail, 75006 Paris, France
hf@ehess.fr, pom@ehess.fr

Summary. We prove that the vertex set of any twin-free loopless multigraph G has an embedding into some point set P of some Euclidean space $\mathbb{R}^k$, such that the automorphism group of G is isomorphic to the isometry group of $\mathbb{R}^k$ globally preserving P.

AMS Subject Classification. Primary 05C50; Secondary 05C62.
Keywords. Automorphism, distance, isometry, spectrum.

1 Introduction

Using spectral analysis, Babai proved in 1978 that the abstract automorphism group of any multigraph G having s distinct eigenvalues with respective multiplicities $m_1, m_2, \ldots, m_s$ is a subgroup of $\boldsymbol{\omega}(m_1) \oplus \boldsymbol{\omega}(m_2) \oplus \ldots \oplus \boldsymbol{\omega}(m_s)$, where $\boldsymbol{\omega}(m)$ denotes the real orthogonal group of dimension m [Bab75]. As a consequence, if all the eigenvalues of G are simple, the only automorphisms of G are involutions.

Some years before, Mani proved that every triconnected planar graph G can be realized as the 1-skeleton of a convex polytope P in $\mathbb{R}^3$ such that all automorphisms of G are induced by isometries of P [Man71]. One non trivial consequence of this result is that the automorphism group $\mathrm{Aut}(G)$ of a triconnected planar graph G has a chain of normal subgroups $\mathrm{Aut}(G) = G_0 \triangleright G_1 \triangleright \ldots \triangleright G_m = 1$ where each quotient G_i/G_{i-1} is either cyclic, or isomorphic to a symmetric group or A_5.

The result of Mani may be expressed in a weaker form: any triconnected planar graph has an embedding f into $\mathbb{R}^3$, such that $\mathrm{Aut}(G)$ is the group of isometries of $\mathbb{R}^3$ globally preserving the point set $P = f(V(G))$, that we shall denote by $\boldsymbol{\omega}(3, P)$.

These two results suggest that a graph G may possibly have some *regular embedding*, that is some embedding $f : V(G) \to \mathbb{R}^k$ such that $\mathrm{Aut}(G)$ is isomorphic to the group $\boldsymbol{\omega}(k, f(V(G)))$ of isometries of $\mathbb{R}^k$ globally preserving

$f(V(G))$, and that this group might be expressed as a subgroup of a group sum relying on spectral considerations. We shall prove that this is indeed the case for any twin-free loopless multigraph. In this paper, multigraphs are always assumed to be loopless.

This result will be proved using techniques similar to the one used in a graph symmetry detection heuristic presented in [Fra99].

In section 2 we recall several definitions and introduce notations. In section 3 we reduce the study of the automorphism groups of multigraphs to the case of *irreducible* multigraphs, that is to multigraphs having no twin vertices. In section 4 we relate embeddings of a multigraphs to metrics and *distance matrices* on its vertex set. We reduce in section 5 the problem of finding regular embeddings to the one of finding metrics on the vertex set of the multigraph which define *Euclidean, reconstructing* and *commuting* distance matrices. We prove in section 6 that Euclidean distance matrices may be built from symmetric real matrices with 0 on the diagonal, what we call *predistance matrices*. The built distance matrix is commuting/reconstructing if the predistance is. We deduce in section 7 that any commuting reconstructing predistance matrix defines a regular embedding and give some examples of reconstructing commuting predistance matrices. In section 8 we study the special case of regular multigraphs and give a strengthened version of Babai's result [Bab75] in this context. Section 9 is devoted to concluding remarks.

2 Definitions and Notations

Let G be a multigraph. For any $x, y \in V(G)$, the *multiplicity* $\mu(x, y)$ is the number of edges (possibly 0) having x and y as endpoints. An *automorphism* of G is a one-to-one mapping $g : V(G) \to V(G)$ such that $\mu(g(x), g(y)) = \mu(x, y)$ for any $x, y \in V(G)$. The automorphisms of G define the *automorphism group* $\mathrm{Aut}(G)$ of G.

In order to ease the matrix presentation, we shall assume that the vertex set of a multigraph G of order n is $\{1, \ldots, n\}$. Then the *adjacency matrix* of G is the symmetric matrix with entries $\boldsymbol{A}_{i,j} = \mu(i, j)$ and any automorphism $g \in \mathrm{Aut}(G)$ may be written as a permutation matrix $\boldsymbol{g}$, where $\boldsymbol{g}_{i,j} = \delta_{j,g(i)}$ (δ is the usual Kronecker symbol). By permissive abridgment, $\mathrm{Aut}(G)$ would as well denote the group of these permutation matrices. Notice that $\mathrm{Aut}(G)$ may be then described as the group of the permutation matrices commuting with $\boldsymbol{A}$. Also, we will denote by $\boldsymbol{E_i}$ the n column matrix whose jth entry is $\delta_{i,j}$, $\mathbf{1}$ the n column matrix filled with 1s and by $\boldsymbol{J}$ the $n \times n$ matrix $\mathbf{1}\mathbf{1}^{\mathrm{T}}$.

The symmetric group acting on a finite set P will be denoted by $\mathfrak{S}(P)$. The real orthogonal group of dimension n will be denoted by $\omega(n)$. If P is a set of points of $\mathbb{R}^n$, the subgroup of $\omega(n)$ globally preserving P will be denoted by $\omega(n, P)$. Notice that $\omega(n, P) < \mathfrak{S}(P)$.

3 Reducing Graphs

Definition 3.1. *Let G be a multigraph. G is* reducible *if there exists two vertices x_1, x_2 such that $\mu(x_1, y) = \mu(x_2, y)$ for any $y \in V(G) \setminus \{x_1, x_2\}$. The vertices x_1 and x_2 are said to be* twins. *If a multigraph is twin-free, it is* irreducible.

Lemma 3.2. *Let X be a subset of vertices, any two elements of which are twins. Then, $\mu(x, y)$ is constant for $x \neq y \in X$.*

Proof. Let $X = \{x_1, x_2, \ldots, x_k\}$ Then, for $1 < i < j \leq k$: $\mu(x_i, x_j) = \mu(x_1, x_j)$ as x_1 and x_i are twins and $\mu(x_1, x_j) = \mu(x_1, x_2)$ as x_2 and x_j are twins. ◻

Lemma 3.3. *Let G be a multigraph. For any distinct vertices x, y, z the following holds:*

$$\begin{aligned} &\text{If } x \text{ and } y \text{ are twins, that is: } && \forall v \in V(G) \setminus \{x, y\},\ \mu(x, v) = \mu(y, v) \\ &\text{and } y \text{ and } z \text{ are twins, that is: } && \forall v \in V(G) \setminus \{y, z\},\ \mu(y, v) = \mu(z, v) \\ &\text{then } x \text{ and } z \text{ are twins, that is: } && \forall v \in V(G) \setminus \{x, z\},\ \mu(x, v) = \mu(z, v). \end{aligned}$$

Proof. Let $v \in V(G) \setminus \{x, z\}$.

If $v \neq y$: $\mu(x, v) = \mu(y, v)$ (as x and y are twins) and $\mu(y, v) = \mu(z, v)$ (as y and z are twins) hence $\mu(x, v) = \mu(z, v)$.

Otherwise $\mu(x, z) = \mu(x, y)$ (as y and z are twins) and $\mu(x, z) = \mu(y, z)$ (as x and y are twins) hence $\mu(x, v) = \mu(z, v)$. ◻

Corollary 3.4. *The vertex set of any multigraph G has a unique partition $\mathcal{P} = \{V_1, \ldots, V_k\}$, such that any V_i is a maximal subset of twin vertices of G. This partition is the* twin-decomposition *of G.*

Corollary 3.5. *Let G be a multigraph, let $\mathcal{P}$ be its twin decomposition and let $G/\mathcal{P}$ denote the quotient multigraph. Then*

$$\mathrm{Aut}(G) = \bigoplus_{X \in \mathcal{P}} \mathfrak{S}(X) \oplus \mathrm{Aut}(G/\mathcal{P}) \tag{1}$$

Moreover, no subset $Y \subseteq V(G/\mathcal{P})$ of cardinality at least two is such that $\mathfrak{S}(Y)$ is a subgroup of $\mathrm{Aut}(G/\mathcal{P})$.

Proof. Assume there exists a subset $Y \subseteq V(G/\mathcal{P})$ of cardinality at least two is such that $\mathfrak{S}(Y)$ is a subgroup of $\mathrm{Aut}(G/\mathcal{P})$. Let τ be a transposition in $\mathfrak{S}(Y)$ exchanging two vertices a and b of $G/\mathcal{P}$. As τ is an automorphism of $G/\mathcal{P}$ it follows that a and b are twins in $G/\mathcal{P}$. Identifying a and b with the classes of twins of vertices in G, it follows that any $x \in a$ is a twin of any $y \in b$ in G hence $a \cup b$ is a class of twins, which contradicts the maximality of the classes in $\mathcal{P}$. ◻

4 Embeddings and Distances

Recall that a *metric* on a set X is a mapping $\mathrm{d} : X^2 \to \mathbb{R}^+$ satisfying the usual axioms of a metric, that is:

$$\begin{aligned} \forall (x,y) \in X^2, &\qquad \mathrm{d}(x,y) = \mathrm{d}(y,x) \\ \forall (x,y) \in X^2, &\qquad \mathrm{d}(x,y) = 0 \iff x = y \\ \forall (x,y,z) \in X^3, &\qquad \mathrm{d}(x,z) \leq \mathrm{d}(x,y) + \mathrm{d}(y,z) \end{aligned}$$

An *Euclidean metric* on a set X is a metric d on X such that there exist some Euclidean space $\mathbb{R}^k$ and an embedding $f : X \to \mathbb{R}^k$ so that, for any $x, y \in X$:

$$\mathrm{dist}(f(x), f(y)) = \mathrm{d}(x,y)$$

where dist is the Euclidean metric of $\mathbb{R}^k$. Any embedding of a multigraph G into some Euclidean space $\mathbb{R}^k$ defines an Euclidean metric d_f on $V(G)$ by $\mathrm{d}_f(x,y) = \mathrm{dist}(f(x), f(y))$.

Definition 4.1. *The* distance matrix *of a metric* d *on the set* $\{1, \dots, n\}$ *is the* $n \times n$ *real symmetric matrix* $\boldsymbol{D}$ *defined by:*

$$\forall i,j \in \{1,\dots,n\}, \qquad \boldsymbol{D}_{i,j} = \mathrm{d}(i,j)^2$$

Notice that the entries of $\boldsymbol{D}$ are the **squares** of the distances and not the distances themselves.

Definition 4.2. *Let* G *be a multigraph and let* $\boldsymbol{D}$ *be a distance matrix of a metric defined on* $V(G) = \{1, \dots, n\}$.

The distance matrix $\boldsymbol{D}$ *is*

- Euclidean *if the metric from which* $\boldsymbol{D}$ *comes from is Euclidean. A compatible embedding of* G *into an Euclidean space* $\mathbb{R}^k$ *is then called a* $\boldsymbol{D}$-embedding *of* G*;*
- reconstructing *if* G *may be reconstructed from* $\boldsymbol{D}$*, that is:*

$$\exists \Xi : \mathbb{R} \to \mathbb{N}, \quad \forall i \neq j \in \{1,\dots,n\}, \quad \mu(i,j) = \Xi(\boldsymbol{D}_{i,j});$$

- commuting *if any automorphism of* G *commutes with* $\boldsymbol{D}$*, that is:*

$$\forall g \in \mathrm{Aut}(G), \qquad \boldsymbol{g}\boldsymbol{D} = \boldsymbol{D}\boldsymbol{g}$$

or, equivalently:

$$\forall g \in \mathrm{Aut}(G), \forall i,j \in \{1,\dots,n\} \qquad \boldsymbol{D}_{g(i),g(j)} = \boldsymbol{D}_{i,j}.$$

5 Distance Matrices and Regular Embeddings

Definition 5.1. *Let G be a multigraph and let k be an integer. A* regular embedding *of G into $\mathbb{R}^k$ is a mapping $f : V(G) \to \mathbb{R}^k$ such that* Aut(G) *is isomorphic to $\omega(k, f(V(G)))$.*

We recall the classical following theorem (see for instance [WW75]):

Theorem 5.2 (Isometry Extension Theorem). *Let $(a_i)_{i\in I}$ and $(b_i)_{i\in I}$ be families of points of an Euclidean affine space $\mathcal{E}$ (with distance* dist*) such that:*

$$\forall (i,j) \in I^2 \qquad \mathrm{dist}(a_i, a_j) = \mathrm{dist}(b_i, b_j)$$

Then there exists an isometry f of $\mathcal{E}$ such that:

$$\forall i \in I \qquad f(a_i) = b_i$$

Moreover, if $(b_i)_{i\in I}$ spans $\mathcal{E}$ then f is uniquely determined.

Theorem 5.3. *Let f be a one-to-one embedding of a multigraph G into the Euclidean space $\mathbb{R}^k$ (with metric* dist*) and let $\boldsymbol{D}$ be the Euclidean distance matrix of the metric defined by f on $V(G)$.*

Assume $\boldsymbol{D}$ is reconstructing. Then f is regular if and only if $f(V(G))$ spans $\mathbb{R}^k$ and $\boldsymbol{D}$ is commuting.

Proof. Assume $f(V(G))$ spans $\mathbb{R}^k$ and that $\boldsymbol{D}$ is commuting.

Let dist be the usual Euclidean metric of $\mathbb{R}^k$. Let $g \in \mathrm{Aut}(G)$. Then, for any vertices x, y of G, $\mathrm{dist}(f(g(x)), f(g(y))) = \mathrm{dist}(f(x), f(y))$. According to Theorem 5.2, g may be extended to a unique isometry of $\mathbb{R}^k$ preserving $f(V(G))$, that is: to a unique element of $\omega(k, f(V(G))$.

Let $\tilde{g}$ denote the isometry defined by the automorphism g. Obviously, if $g_1, g_2 \in \mathrm{Aut}(G)$, $\widetilde{g_1 g_2} = \tilde{g_1}\tilde{g_2}$. Thus we have defined a group morphism from $\mathrm{Aut}(G)$ to $\omega(k, f(V(G)))$.

Now assume $\phi \in \omega(k, f(V(G)))$. Define $g : V(G) \to V(G)$ by $f(g(x)) = \phi(f(x))$. Notice that g is well defined as ϕ is obviously one-to-one. As $\boldsymbol{D}$ is reconstructing, there exists a mapping $\Xi : \mathbb{R} \to \mathbb{N}$ such that $\Xi(\boldsymbol{D}_{i,j}) = \mu(i,j)$. Let $i \neq j \in \{1, \ldots, n\}$. Then, for any $i \neq j \in \{1, \ldots, n\}$, we have: $\mu(i,j) = \Xi(\boldsymbol{D}_{i,j})$. As ϕ is an isometry, the distance between $f(i)$ and $f(i)$ is the same as the distance as the distance between $\phi(f(i))$ and $\phi(f(i))$, that is: between $f(g(i))$ and $f(g(i))$ (by the definition of g). It follows that $\boldsymbol{D}_{g(i),g(j)} = \boldsymbol{D}_{i,j}$ hence $\mu(g(i), g(j)) = \mu(i,j)$. It follows that g is an automorphism of G. Moreover, ϕ extends $f^{-1} \circ g \circ f$, thus $\phi = \tilde{g}$. It follows that $g \mapsto \tilde{g}$ is actually a group isomorphism from $\mathrm{Aut}(G)$ to $\omega(k, f(V(G)))$.

Conversely, assume f is a regular embedding. Then $f(V(G))$ spans $\mathbb{R}^k$ for otherwise $\omega(k, f(V(G)))$ would not be a finite group although $\mathrm{Aut}(G)$ is. Let $\phi \in \omega(k, f(V(G)))$. Define $\tilde{\phi} : V(G) \to V(G)$ by $f(\tilde{\phi}(x)) = \phi(f(x))$. As previously, we get $\mu(\tilde{\phi}(x), \tilde{\phi}(y)) = m$, hence $\tilde{\phi}$ is an automorphism of G such that $\mathrm{d}(f(\tilde{\phi}(x)), f(\tilde{\phi}(y))) = \mathrm{d}(f(x), f(y))$. Moreover, $\widetilde{\phi_1\phi_2} = \tilde{\phi}_1\tilde{\phi}_2$. It follows that $\phi \mapsto \tilde{\phi}$ is a group-morphism. As it is clearly one-to-one and as $\mathrm{Aut}(G)$ and $\omega(k, f(V(G)))$ are isomorphic by assumption, $\phi \mapsto \tilde{\phi}$ is onto. It follows that $\mathrm{d}(f(g(x)), f(g(y))) = \mathrm{d}(f(x), f(y))$ for any $g \in \mathrm{Aut}(G)$, that is: $\boldsymbol{D}$ is commuting. □

We shall consider now this result from another point of view:

Corollary 5.4. *Let G be a multigraph and let $\boldsymbol{D}$ be an Euclidean reconstructing and commuting distance matrix on G. Then $\boldsymbol{D}$ defines a regular embedding of G.*

Proof. Consider any $\boldsymbol{D}$-embedding f of G and the subspace spanned by $f(V(G))$ and apply Theorem 5.3. □

Our main problem is now to build such a distance matrix.

6 Euclidean Distance Matrices from Predistance Matrices

Given a distance dist on the set $\{1, \ldots, n\}$, it is classical to define the corresponding bilinear form $\boldsymbol{B}$ by:

$$\boldsymbol{B} = -\frac{1}{2}(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{D}(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J}) \tag{2}$$

where $\boldsymbol{D}_{i,j} = \mathrm{dist}(i,j)^2$.

It is well known that the distance dist is Euclidean (i.e. allows an isometric embedding into some Euclidean space) if and only if $\boldsymbol{B}$ is positive semi-definite [B&73, HY64]. Such a result extends to a characterization of those symmetric real matrices which are Euclidean distance matrices.

Definition 6.1. *A* predistance matrix *on G is an $n \times n$ symmetric real matrix $\boldsymbol{P}$ with 0 on the diagonal (notice that negative entries are allowed).*

The bilinear form $\Lambda(\boldsymbol{P})$ of $\boldsymbol{P}$ *is defined by:*

$$\Lambda(\boldsymbol{P}) = -\frac{1}{2}(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})\boldsymbol{P}(\boldsymbol{I} - \frac{1}{n}\boldsymbol{J})$$

Lemma 6.2. *Let $\boldsymbol{D}$ be a predistance matrix and let k be an integer. Then, the following statements are equivalent:*

1. *$\boldsymbol{D}$ is an Euclidean distance matrix on G, and there exists a D-embedding of G in $\mathbb{R}^k$;*
2. *the matrix $\Lambda(\boldsymbol{D})$ is positive semi-definite and* $\mathrm{rank}(\Lambda(\boldsymbol{D})) \leq k$.

Proof. Assume $\Lambda(\boldsymbol{D})$ is positive semi-definite. Then the square of the distance defined by $\Lambda(\boldsymbol{D})$ between the basis points e_i and e_j is $(\boldsymbol{E}_i-\boldsymbol{E}_j)^{\mathrm{T}}\Lambda(\boldsymbol{D})(\boldsymbol{E}_i-\boldsymbol{E}_j)$. As $\left(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\right)(\boldsymbol{E}_i-\boldsymbol{E}_j)=(\boldsymbol{E}_i-\boldsymbol{E}_j)$, we get that the square of this distance equals $(\boldsymbol{D}_{i,j}+\boldsymbol{D}_{j,i}-\boldsymbol{D}_{i,i}-\boldsymbol{D}_{j,j})/2=\boldsymbol{D}_{i,j}$ as $\boldsymbol{D}$ is symmetric with zero on the diagonal. We define $f(i)$ as the projection of e_i orthogonal to the isotropic space of B.

Conversely, assume $\boldsymbol{D}$ is an Euclidean distance matrix and let $f:[\,1\,;n\,]\to\mathbb{R}^k$ be a $\boldsymbol{D}$-embedding of G. Then $\Lambda(\boldsymbol{D})$ is the Gram matrix of the vectors $f(i)$, i.e. $\Lambda(\boldsymbol{D})_{i,j}=<f(i),f(j)>$. It is well known that this Gram matrix determines the vectors $f(i)$ up to isometry. ◻

Let $\boldsymbol{P}$ be a predistance matrix. The condition that $\Lambda(\boldsymbol{P})$ has no negative eigenvalue is quite difficult to handle. However, we know the following about $\Lambda(\boldsymbol{P})$:

- $\Lambda(\boldsymbol{P})$ is symmetric real and thus diagonalizable,
- $\mathbf{1}\in\ker(\Lambda(\boldsymbol{P}))$ as $(\boldsymbol{I}-\frac{1}{n}\boldsymbol{J})\mathbf{1}=\mathbf{0}$,
- $\Lambda(\boldsymbol{P})$ has an orthogonal basis $\mathcal{B}(\Lambda(\boldsymbol{P}))$ of eigenvectors including $\mathbf{1}$. The eigenvectors of $\mathcal{B}(\Lambda(\boldsymbol{P}))\setminus\{\mathbf{1}\}$ have eigenvalues $\lambda_1>\ldots>\lambda_r$ with respective multiplicities $m_1,\ldots,m_r$.

Now consider the *reduced distance matrix* $\boldsymbol{P}^\star=\boldsymbol{P}-2\lambda_r(\boldsymbol{J}-\boldsymbol{I})$.

Lemma 6.3. *$\mathcal{B}(\Lambda(\boldsymbol{P}))$ is an orthogonal basis of eigenvectors of $\Lambda(\boldsymbol{P}^\star)$, and the eigenvalue (for $\Lambda(\boldsymbol{P}^\star)$) of $v\in\mathcal{B}(\Lambda(\boldsymbol{P}))$ is*

$$\begin{cases} 0 & \text{if } \boldsymbol{v}=\mathbf{1}\\ \lambda-\lambda_r & \text{if } \boldsymbol{v}\neq\mathbf{1} \text{ and } \Lambda(\boldsymbol{P})\boldsymbol{v}=\lambda\boldsymbol{v}\end{cases}$$

Thus, $\Lambda(\boldsymbol{P}^\star)$ is positive semi-definite and has corank m_r+1.

Proof. If $\boldsymbol{v}\neq\mathbf{1}$, then $\boldsymbol{v}\in\mathbf{1}^\perp$, and thus $\left(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\right)\boldsymbol{v}=\boldsymbol{v}$. Hence, if $\Lambda(\boldsymbol{P})\boldsymbol{v}=\lambda\boldsymbol{v}$, then

$$\begin{aligned}
\Lambda(\boldsymbol{P}^\star)\boldsymbol{v} &= -\frac{1}{2}\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)\boldsymbol{P}^\star\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)\boldsymbol{v}\\
&= -\frac{1}{2}\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)(\boldsymbol{P}-2\lambda_r(\boldsymbol{J}-\boldsymbol{I}))\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)\boldsymbol{v}\\
&= -\frac{1}{2}\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)\boldsymbol{P}\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)\boldsymbol{v}+\lambda_r\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)(\boldsymbol{J}-\boldsymbol{I})\big(\boldsymbol{I}-\frac{\boldsymbol{J}}{n}\big)\boldsymbol{v}\\
&= \Lambda(\boldsymbol{P})\boldsymbol{v}-\lambda_r\boldsymbol{v}(\text{as } \boldsymbol{J}\boldsymbol{v}=\mathbf{0})\\
&= (\lambda-\lambda_r)\boldsymbol{v}
\end{aligned}$$

◻

7 Reconstructing and Commuting Predistance Matrices

Definition 7.1. *A predistance matrix* $\boldsymbol{P}$ *is*

- commuting *if* $\boldsymbol{P}$ *commutes with any automorphism of* G*, that is:*

$$\forall g \in \mathrm{Aut}(G), \qquad \boldsymbol{gP} = \boldsymbol{Pg}$$

  or, equivalently:

$$i, j \in \{1, \ldots, n\}, \qquad \boldsymbol{P}_{g(i),g(j)} = \boldsymbol{P}_{i,j};$$

- reconstructing *if there exists a mapping* $\Xi : \mathbb{R} \to \mathbb{N}$ *such that* $\mu(i,j) = \Xi(\boldsymbol{P}_{i,j})$*, or equivalently:*

$$\forall i, j, i', j' \in \{1, \ldots, n\} \qquad \boldsymbol{P}_{i,j} = \boldsymbol{P}_{i',j'} \Rightarrow \mu(i,j) = \mu(i',j')$$

Notation 7.2. Let $\boldsymbol{P}$ be a predistance matrix. As $\Lambda(\boldsymbol{P})$ has 0 as an eigenvalue, we may define $\zeta(\boldsymbol{P})$ as follows:

- if the smallest eigenvalue λ of $\Lambda(\boldsymbol{P})$ is negative and has multiplicity m then $\zeta(\boldsymbol{P}) = n - m - 1$,
- if the smallest eigenvalue of $\Lambda(\boldsymbol{P})$ is 0 with multiplicity $m > 1$ then $\zeta(\boldsymbol{P}) = n - m$,
- if the smallest eigenvalue of $\Lambda(\boldsymbol{P})$ is 0 with multiplicity 1 and if the second smallest eigenvalue of $\Lambda(\boldsymbol{P})$ has multiplicity m then $\zeta(\boldsymbol{P}) = n - m - 1$.

Theorem 7.3. *Let* G *be an irreducible multigraph. Any commuting reconstructing predistance matrix* $\boldsymbol{P}$ *defines a regular embedding* f *into* $\mathbb{R}^{\zeta(\boldsymbol{P})}$.

Proof. Let $\boldsymbol{P}$ be a commuting reconstructing predistance matrix $\boldsymbol{P}$. The corresponding reduced distance matrix $\boldsymbol{P}^\star$, according to Lemma 6.2 and Lemma 6.3 is Euclidean and there exists a $\boldsymbol{P}^\star$ embedding f on G into $\mathbb{R}^{\zeta(\boldsymbol{P})}$. As $\boldsymbol{P}$ is reconstructing and commuting, so is $\boldsymbol{P}^\star$. Assume $f(x) = f(y)$. Since f is reconstructing, x and y are twins of G, contradicting its irreducibility. Hence f is one-to-one and, according to Theorem 5.3, f is a regular embedding. ◻

Corollary 7.4. *Let* G *be an irreducible multigraph and let* $\boldsymbol{P}$ *be a reconstructing predistance matrix which commutes with the automorphisms of* G*.*

Denote by $\mathbf{1}\mathbb{R}$ *the line spanned by* $\mathbf{1}$ *and by* $\overset{\perp}{\oplus}$ *the orthogonal direct sum of vector spaces. Let* $\mathbf{1}\mathbb{R} \overset{\perp}{\oplus} E_1 \overset{\perp}{\oplus} E_2 \overset{\perp}{\oplus} \ldots \overset{\perp}{\oplus} E_r$ *be a decomposition into eigenspaces of* $\Lambda(\boldsymbol{P})$*, the eigenvalues associated with* $E_1, \ldots, E_r$ *being* $\lambda_1 > \lambda_2 > \ldots > \lambda_r$ *(notice that some* λ_i *may be* 0*). Then the abstract automorphism group of* G *is a subgroup of*

$$\boldsymbol{\omega}(\dim E_1, P_1) \oplus \boldsymbol{\omega}(\dim E_2, P_2) \oplus \cdots \oplus \boldsymbol{\omega}(\dim E_{r-1}, P_{r-1})$$

where P_i *is the orthogonal projection of the image of* G *under the regular embedding defined by* $\boldsymbol{P}$ *into the subspace* E_i.

Here are some examples of simple commuting reconstructing predistance matrices of a simple connected graph G (the two last examples only apply if G has order at least 3):

Adjacency $$\boldsymbol{A}_{i,j} = \begin{cases} 0, & \text{if } i = j \text{ or } i \text{ and } j \text{ are adjacent} \\ 1, & \text{otherwise} \end{cases}$$

Graph distance $$\boldsymbol{S}_{i,j} = \begin{cases} 0, & \text{if } i = j \\ l, & \text{if } l \text{ is the graph distance from } i \text{ to } j \end{cases}$$

Bisected Czekanovski-Dice $$\boldsymbol{C}_{i,j} = \begin{cases} 0, & \text{if } i = j \\ 1 - \frac{2}{d(i)+d(j)}, & \text{if } i \text{ and } j \text{ are adjacent} \\ 1, & \text{otherwise} \end{cases}$$

Q-distance $$\boldsymbol{Q}_{i,j} = \begin{cases} 0, & \text{if } i = j \\ 1, & \text{if } i \text{ and } j \text{ are non adjacent} \\ 1 - \frac{1}{\sqrt{d(i)d(j)}}, & \text{otherwise} \end{cases}$$

Theorem 7.5. *Any irreducible multigraph has a regular embedding into some Euclidean space.*

Proof. The adjacency matrix $\boldsymbol{A}$ of G defined by $\boldsymbol{A}_{i,j} = \mu(i,j)$ is commuting and reconstructing. Hence the result follows from Theorem 7.3. ◻

8 Regular Multigraphs

As noted in section 7, the adjacency matrix $\boldsymbol{A}$ defines a commuting reconstructing predistance matrix. In the particular case where G is d-regular, we have $\boldsymbol{AJ} = d\boldsymbol{J}$ thus $\Lambda(\boldsymbol{A}) = (1 - d/n)^2\boldsymbol{A}$. Moreover (see [Big74]):

- d is an eigenvalue of G with eigenvector $\mathbf{1}$.
- If G is connected, then the multiplicity of d is one.
- For any eigenvalue λ of G, we have $|\lambda| \leq d$.

In this case we deduce a strengthening of Babai's result [Bab75] mentioned in section 1:

Corollary 8.1. *Let G be a connected irreducible regular multigraph having s distinct eigenvalues $\lambda_1 > \lambda_2 > \ldots > \lambda_s$ with respective multiplicities $m_1, m_2, \ldots, m_s$. Then, the abstract automorphism group γ of G is a subgroup of*

$$\boldsymbol{\omega}(m_2, P_2) \oplus \boldsymbol{\omega}(m_3, P_3) \oplus \ldots \oplus \boldsymbol{\omega}(m_{s-1}, P_{s-1}).$$

Proof. Assume G is connected. Let $\mathcal{B}(\boldsymbol{A})$ be an orthogonal basis of eigenvectors of $\boldsymbol{A}$. As $\boldsymbol{Jv} = 0$ for any vector orthogonal to $\mathbf{1}$ and as $\boldsymbol{J}\mathbf{1} = n\mathbf{1}$, $\mathcal{B}(\boldsymbol{A})$ is also a basis of eigenvectors of $\Lambda(\boldsymbol{A})$. The eigenvalue (for $\Lambda(\boldsymbol{A})$) of an eigenvector $\boldsymbol{v}$ with eigenvalue λ (for $\boldsymbol{A}$) is clearly λ if $\lambda \neq d$ and 0 if $\lambda = d$. As the eigenvalue d is maximal and simple for $\boldsymbol{A}$, we deduce that $\zeta(\boldsymbol{A}) = n - m - 1$, where m is the multiplicity of the smallest eigenvalue of $\boldsymbol{A}$. Moreover, the greatest eigenvalue is simple and corresponds to eigenvector $\mathbf{1}$, according to Perron-Frobenius theorem [Fro12, Per07]. □

As a matter of fact, instead of $\boldsymbol{\omega}(m_2, P_2) \oplus \boldsymbol{\omega}(m_3, P_3) \oplus \ldots \oplus \boldsymbol{\omega}(m_{s-1}, P_{s-1})$, we may prove a similar result with $\boldsymbol{\omega}(m_3, P_3) \oplus \ldots \oplus \boldsymbol{\omega}(m_{s-1}, P_{s-1}) \oplus \boldsymbol{\omega}(m_s, P_s)$ by considering $2\lambda_2(\boldsymbol{J} - \boldsymbol{I}) - \boldsymbol{A}$ instead of $\boldsymbol{A} - 2\lambda_s(\boldsymbol{J} - \boldsymbol{I})$.

A *strongly regular graph* is a regular graph such that there exist constants λ, μ so that any two adjacent vertices have exactly λ common neighbors, and every two non-adjacent vertices have exactly μ common neighbors [Sei79]. As a strongly regular graph has exactly three distinct eigenvalues [Big74], we may choose to keep among E_2 and E_3 the one having the smallest dimension. Hence every connected irreducible strong regular graph of order n has a regular embedding into an Euclidean space of dimension at most $\lfloor \frac{n-1}{2} \rfloor$. As an example, the Petersen graph has a regular embedding into $\mathbb{R}^4$, which is optimal (the automorphism group of the Petersen graph cannot be realized as the isometry group of a set of points in $\mathbb{R}^3$).

9 Conclusion

The techniques presented here allow to construct regular embeddings from weakly constrained matrices: we only ask that they should be symmetric, have 0 on the diagonal, are reconstructing and commute with every automorphism.

Recall that a *cellular algebra*, or *coherent algebra*, is an algebra of $n \times n$ complex matrices which has a basis $\{\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_t\}$ consisting of matrices with entries 0 and 1 satisfying the following conditions:

1. $\boldsymbol{B}_0 + \boldsymbol{B}_1 + \cdots + \boldsymbol{B}_t = \boldsymbol{J}$;
2. $\exists 0 \leq r \leq t : \quad \boldsymbol{B}_0 + \boldsymbol{B}_1 + \cdots + \boldsymbol{B}_r = \boldsymbol{I}$;
3. the set $\{\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_t\}$ is closed under transposition.

The unique minimal cellular algebra which contains $\boldsymbol{A}$ as an element is the *cellular algebra generated by* $\boldsymbol{A}$ and is denoted by $<< \boldsymbol{A} >>$. Notice that a basis of this cellular algebra may be constructed in polynomial time. In the case where $\boldsymbol{A}$ is the adjacency matrix of a simple graph G, the cellular algebra $<< \boldsymbol{A} >>$ is called the cellular algebra *generated by* G, or the cellular algebra *of* G. This algebra may contain non symmetric basis elements. However, any matrix in $<< A >>$ commutes with any automorphism of G. It follows that any commuting predistance matrix $\boldsymbol{P}$ in $<< A >>$ may be written as $\sum_{i=r+1}^{t} \lambda_i(\boldsymbol{B}_i + \boldsymbol{B}_i^{\mathrm{T}})$, with $\lambda_i \in \mathbb{R}$ for $r < i \leq t$ and thus form an affine space $\mathcal{E}$. By perturbing each of them (by adding $\epsilon(\boldsymbol{P})\boldsymbol{A}$, for instance) we obtain reconstructing and commuting predistance matrices forming a dense subset of $\mathcal{E}$.

We hope that the suggested approach will be fruitful and more practical than the usual techniques arising from algebra and spectral analysis.

References

[Bab75] L. Babai, Automorphism group and category of cospectral graphs, *Acta Math. Acad. Sci. Hung.* **31** (1978), 195–306.

[B&73] J. P. Benzécri and et al, *L'analyse des données — Tome II: l'analyse des correspondances*, Dunod, Paris, 1973.

[Big74] N. L. Biggs, *Algebraic graph theory*, Cambridge tracts in mathematics, vol. 67, Cambridge University Press, 1974.

[Fra99] H. de Fraysseix, An heuristic for graph symmetry detection, *Graph Drawing* (J. Kratochvíl, ed.), *Lecture Notes in Computer Science*, vol. 1731, Springer, 1999, pp. 276–285.

[Fro12] G. Frobenius, Über Matrizen aus nicht negativen Elementen, S.-B. Preuss. Akad. Wiss. Berlin (1912), 456–477.

[HY64] S. L. Hakimi and S. S. Yau, Distance matrix of a graph and its realizability, *Quaterly of Applied Mathematics* **22** (1964), 305–317.

[Man71] P. Mani, Automorphismen von polyedrischen Graphen, *Math. Ann.* **192** (1971), 279–303.

[Per07] O. Perron, Zur Theorie der Matrizen, *Math. Ann.* **64** (1907), 248–263.

[Sei79] J. J. Seidel, Strongly regular graphs, In *Surveys in Combinatorics (Proc. Seventh British Combinatorial Conf., Cambridge, 1979)* (Cambridge University Press, ed.), 1979, pp. 157–180.

[WW75] J. H. Wells and L. R. Williams, *Embeddings and extensions in analysis*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 84.e, Springer-Verlag, 1975.

Quadrangulations and 5-critical Graphs on the Projective Plane

Bojan Mohar*

Department of Mathematics, University of Ljubljana, Ljubljana, Slovenia
and Department of Mathematics, Simon Fraser University, Burnaby, B.C.
mohar@sfu.ca

Summary. Let Q be a nonbipartite quadrangulation of the projective plane. Youngs [You96] proved that Q cannot be 3-colored. We prove that for every 4-coloring of Q and for any two colors a and b, the number of faces F of Q, on which all four colors appear and colors a and b are not adjacent on F, is odd. This strengthens previous results that have appeared in [You96, HRS02, Moh02, CT04]. If we form a triangulation of the projective plane by inserting a vertex of degree 4 in every face of Q, we obtain an Eulerian triangulation T of the projective plane whose chromatic number is 5. The above result shows that T is never 5-critical. We show that sometimes one can remove two, three, or four, vertices from T and obtain a 5-critical graph. This gives rise to an explicit construction of 5-critical graphs on the projective plane and yields the first explicit family of 5-critical graphs with arbitrarily large edge-width on a fixed surface.

AMS Subject Classification. 05C10, 05C15.
Keywords. Quadrangulation, Eulerian triangulation, coloring, projective plane, 5-critical graph, edge-width.

1 Introduction

Youngs [You96] proved that a quadrangulation of the projective plane which is not bipartite is never 3-colorable, and its chromatic number is 4. Youngs' proof also implies that in any 4-coloring of a nonbipartite quadrangulation Q of the projective plane, there is a 4-face with all four vertices of distinct colors. This fact appears in a slightly extended version (where 4-colorings are replaced by k-colorings, $k \geq 3$) in [HRS02]. A strengthening of that result, proved in [Moh02], states that under every k-coloring of Q, there are at least three faces on which all four vertices have distinct colors. This in particular implies that $k \geq 4$. Collins and Tysdal [CT04] found that every 4-coloring of Q has a

* Supported in part by the Ministry of Higher Education, Science and Technology of Slovenia, Research Project J1-0502 and Research Program P1-0297.

face F on which all four colors appear and in which colors 1 and 2 are used on opposite vertices of F. In this paper we give further extensions of these results in Theorems 2.1 and 3.1.

Let G be a graph and k a positive integer. We say that G is *k-critical* if its chromatic number is k, but every proper subgraph of G has chromatic number at most $k-1$.

For a fixed surface S and every $k \geq 8$, there are only finitely many k-critical graphs that can be embedded in S. This follows easily from Euler's formula and the fact that a k-critical graph cannot have vertices of degree smaller than $k-1$. Slightly more involved arguments imply that the same holds for 7-critical graphs on a fixed surface. See [Edw92, Moh93] or [MT01, Section 8.4]. Thomassen [Tho97] extended this to 6-critical graphs (and the proofs became much more complicated at this point). A corollary of these results is that for every surface S, there is a polynomial time algorithm to decide if a given graph G embedded in S has chromatic number at least 5, and if so, the algorithm also outputs $\chi(G)$.

On the other hand, 3-coloring planar graphs is NP-hard [GJ79], so it is NP-hard to 3-color graphs on any fixed surface. It is an open problem if 4-coloring on fixed surfaces is polynomially decidable. It is known to be so for planar graphs [RSST96], but unknown for other surfaces.

It is known that every nonplanar surface S contains infinitely many 5-critical graphs. There is a rather "straightforward" argument showing this. The proof goes as follows (see also [MT01, Corollary 8.4.13]). Let T be a triangulation of S such that all vertices of T have even degree except two of them whose degree is odd and are adjacent. Such triangulations are easy to construct on every nonplanar surface. They are called *Fisk triangulations* after a result of Fisk, who proved in [Fisk78] that T cannot be 4-colorable, since every 4-coloring of a triangulation with precisely two vertices of odd degree uses the same color on both vertices of odd degree. For every nonplanar surface S and every integer k, there exists a Fisk triangulation T_k on S whose *edge-width* (the length of a shortest noncontractible cycle) is at least k. Now, let R_k be a 5-critical subgraph of T_k. Since $\chi(R_k) \geq 5$, R_k is nonplanar and hence it contains a cycle that is noncontractible in the induced embedding. That cycle is also noncontractible in T_k and hence it has at least k vertices. In particular, $|R_k| \geq k$, and hence there are infinitely many nonisomorphic graphs in the sequence $R_1, R_2, R_3, \ldots$. The reader may have observed that the above simple argument is not entirely elementary since it uses the 4-color-theorem.

Fisk triangulations are never 5-critical. To see this, let T be a Fisk triangulation, and let x, y be its vertices of odd degree. Let M be the subgraph of T obtained by deleting the two edges xz, yz in a facial triangle containing the edge xy. The new face of M is bounded by a cycle C of length 5, and all vertices of M have even degree. Suppose now that M has a 4-coloring. Then some vertex $u \in V(C)$ has a color which does not appear on other vertices of C. By adding two edges connecting u with the two vertices which are "opposite" on C, we obtain a Fisk triangulation with a 4-coloring, a contradiction.

Let c be a coloring of a graph G. If G is embedded in some surface and F is a face, we say that F is *multicolored* if all vertices of F have distinct colors under the coloring c. The results about nonbipartite quadrangulations of the projective plane mentioned above show that by triangulating every face of such a quadrangulation by inserting a new vertex of degree 4, we obtain a triangulation whose chromatic number is 5, but such a graph is never 5-critical.

In Section 4, we describe some 5-critical subgraphs of triangulated quadrangulations, and we use them to show the following:

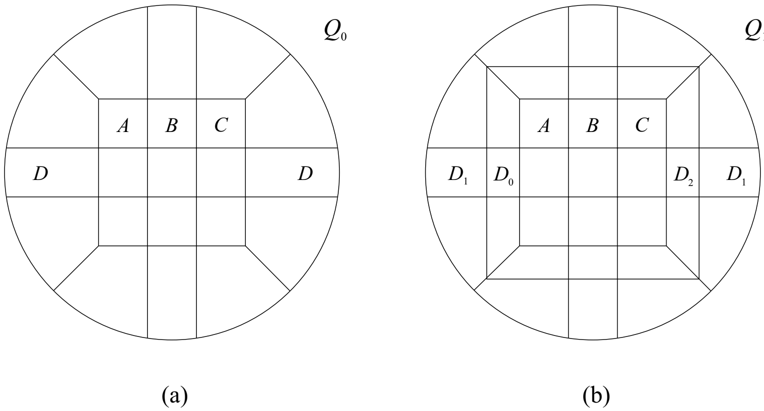


Fig. 1. Projective quadrangulations Q_0 and Q_1

Theorem 1.1. *Let Q_1 be the quadrangulation of the projective plane shown in Figure 1(b), and let A, B, C be the faces of Q_1 as shown in the same figure. Let Q be a quadrangulation of the projective plane that can be obtained by successively replacing a face distinct from A, B, C by one of the graphs K or L shown in Figure 2. Denote by $\mathcal{T}(Q; A, B, C)$ the graph obtained from Q by adding a new vertex in each face F distinct from A, B, C and joining that vertex to all four vertices on the facial walk of F. Then $\mathcal{T}(Q; A, B, C)$ is 5-critical.*

The proof of Theorem 1.1 is given in Section 4.

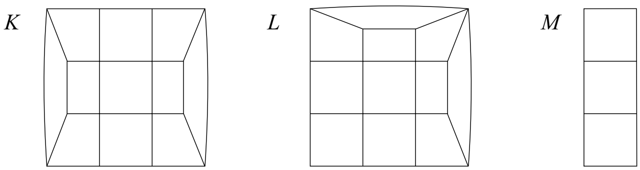


Fig. 2. Subdividers K, L, and M

Of course, replacements of faces of Q_1 by the graphs K and L from Figure 2 has to be done in such a way that a quadrangulation is obtained. In order to achieve this, replacement of a face F by K requires to replace also the faces above and below F (or left and right of F if K is rotated by angle $\pi/2$ before the replacement); similarly for L. After making the first round of replacements, we can repeat the procedure with the faces of the resulting quadrangulation, etc. In this way we can obtain quadrangulations whose edge-width is as large as we want. Henceforth, the family of graphs described in Theorem 1.1 is the first explicitly described family of 5-critical graphs with arbitrarily large edge-width on a fixed surface.

Other constructions of 5-critical graphs on the projective plane are given in Section 4.

Graphs in this paper are finite; multiple edges are allowed, while loops are excluded. If G is a graph, then $|G|$ denotes its order. A *quadrangulation* is a connected graph together with a 2-cell embedding in a closed surface such that every facial walk is of length 4. Note that a pair of edges joining the same pair of vertices can form a contractible curve on the surface of the quadrangulation, but it does not bound a face since all faces are of lenth 4.

2 Quadrangulations of the Sphere

Let Q be a quadrangulation of the sphere. It is easy to see that Q is bipartite and hence 2-colorable. Therefore, it may seem surprising if we would be able to say something about global properties of 4-colorings of Q. In this section, we uncover some surprising properties.

Let c be a k-coloring of an embedded graph. We say that a face F is *multicolored* if all its vertices have distinct colors. Suppose that the coloring uses colors $0, 1, \ldots, k-1$. If $F = xyzw$ is a multicolored 4-face and $c(x) = 0$, then the color $c(z)$ of z is said to be the *type* of F. If the surface is orientable, we may assume that all faces are equipped with the positive orientation. Assuming this and assuming that $k = 4$, we can refine the notion of the type of multicolored 4-faces. We say that the *type* of F is $1^-, 1^+, 2^-, 2^+, 3^-$, or 3^+ if the clockwise cyclic order of colors on F is $0213, 0312, 0123, 0321, 0132$, or 0231, respectively. (So, the “minus” in the type says that $c(y) < c(w)$, and the “plus” says the converse.)

Theorem 2.1. *Let Q be a quadrangulation of the sphere and c its 4-coloring. Then for every type $t \in \{1, 2, 3\}$, the number of multicolored faces of type t^- has the same parity as the number of multicolored faces of type t^+. In particular, the number of multicolored faces of type t is even.*

Proof. The proof is by induction on the number N which is the sum of the number of vertices $n = |Q|$ and the number of multicolored faces of Q. If Q has no multicolored faces, there is nothing to prove. So, we assume that there is at least one multicolored face (and hence we have $n \geq 4$).

If Q has a face $F = xyzw$ in which $c(x) = c(z)$, then we identify x and z (and delete F and replace resulting parallel edges by single edges). This operation gives rise to a smaller 4-colored quadrangulation Q'. We say that Q' has been obtained by *squeezing* F. Clearly, all other faces and their coloring remain the same as in Q, so we just apply the induction hypothesis to Q' and thus complete the proof.

In the next paragraph we will apply an operation for which we need that there are no parallel edges in Q. If there were, we could apply the induction hypothesis, first to the interior of the disk bounded by a couple of parallel edges between vertices x, y, and then to the exterior of the same disk.

We may now assume that all faces of Q are multicolored. Let xy be an edge of Q, and let $F = xyzw$ and $F' = xw'z'y$ be the facial walks containing the edge xy. Supose that $c(z) = c(z')$ (and hence $c(w) = c(w')$), i.e., F and F' are of the same type. In this case we remove faces F and F', identify w with w' and y with y' and keep only single edges between x and w and y and z, respectively. This gives rise to a smaller 4-colored quadrangulation Q' in which the faces have the same types as before. By applying the induction hypothesis and observing that F and F' were of the same type, but with different signs, we easily complete the proof.

From now on we also assume that no adjacent faces are of the same type. Then it is easy to see that every face of type t^+ or t^- is adjacent to four faces of precisely the same four types as shown in Figure 3. This implies that the dual graph of Q is a covering graph of the dual graph of the cube (which is the octahedron graph) shown in the figure. In particular, the number of faces of any of the types t^+ or t^- is equal to the degree of the corresponding covering projection. This completes the proof. □

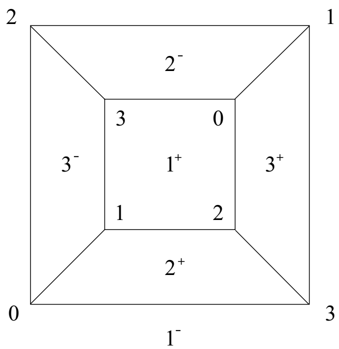


Fig. 3. Types of faces in a 4-colored cube

If Q is a connected plane graph whose interior faces are all 4-sided and the exterior face is of length k, then we say that Q is a *k-near-quadrangulation*. Note that k must be even since Q is bipartite.

Corollary 2.2. *Let Q be a planar 6-near-quadrangulation and let c be a 4-coloring of Q. For $t \in \{1,2,3\}$, let m_t be the number of multicolored 4-faces of Q whose type is t. Then $m_1 \equiv m_2 \equiv m_3 \equiv 1 \pmod 2$ if and only if under the coloring c, every pair of opposite vertices on the exterior face are colored the same.*

Proof. If a pair of opposite vertices x, y have distinct colors, we add the edge xy and obtain a 4-colored quadrangulation. With respect to the coloring of this quadrangulation, at most two of the values m_t can change. Therefore, the one that has not been changed is even by Theorem 2.1. On the other hand, if opposite vertices have the same colors, we quadrangulate the outer face by adding a vertex of degree 3 and color this vertex by the color that does not appear on the exterior face. Then the new faces are all multicolored and of all three types. Again, Theorem 2.1 shows that the values m_1, m_2, and m_3 (in Q) are all odd. This completes the proof. □

We will need another result which will enable us to construct colorings with few multicolored faces.

Lemma 2.3. *Let Q be an 8-near-quadrangulation that is isomorphic to one of K or L shown in Figure 2, and let x, y, z, w be the four corner vertices. Suppose that c_0 is a 4-coloring of x, y, z, w such that $c_0(x) \neq c_0(y) \neq c_0(z) \neq c_0(w) \neq c_0(x)$. For every interior 4-face F of Q, there exists a 4-coloring c of Q which extends c_0 and has the following properties:*

(a) *The vertices on the segment of the outer face from x to y are 2-colored with colors $c_0(x)$ and $c_0(y)$, the vertices from y to z with colors $c_0(y)$ and $c_0(z)$, and similarly for the segments from z to w, and from w to x. There is one exception to this rule if F is the "middle" face having an edge on one of these segments, as shown in Figure 4(d_1) and (d_2) for the segment from x to y. In that case, one of the vertices is not colored as stated, and we may choose either of the two vertices to be this exception. See Figure 4(d_1) and (d_2), where the exceptions are emphasized by little circles.*
(b) *No interior face different from F is multicolored.*
(c) *The colors on F in the clockwise cyclic order are either $c_0(x), c_0(y), c_0(z), c_0(w)$, or the reverse of this.*

Proof. Extensions (up to symmetries) are shown in Figure 4, where a $= c_0(x)$, b $= c_0(y)$, c $= c_0(z)$, and d $= c_0(w)$; the face F is shaded. □

Lemma 2.3 has a generalization which we include for the sake of completeness.

Proposition 2.4. *Let Q be a quadrangulation of the sphere without multiple edges. If F_1, F_2 are distinct faces of Q, then there is a 4-coloring of Q such*

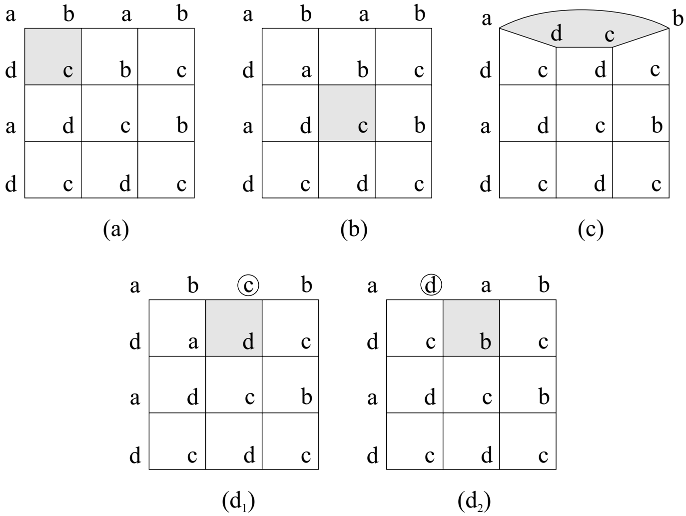


Fig. 4. Extensions of c_0 in K or L

that F_1 and F_2 are multicolored but no other face is multicolored. If the sequence of colors in the clockwise direction on F_1 is $c_1c_2c_3c_4$, then the cyclic sequence of colors on F_2 is its reverse $c_1c_4c_3c_2$. In particular, F_1 and F_2 are of the same type $t \in \{1, 2, 3\}$, but of different signs, i.e., one is of type t^+, the other one of type t^-.

Proof. The proof is by induction on $n = |Q|$. If Q is just a 4-cycle, the statement clearly holds. If Q contains a 4-cycle C that is not facial, let Q_1 be its interior. We may assume that F_1 is a face of Q_1. Let Q_2 be the exterior of C. Now, C is facial in Q_1 and in Q_2. If F_2 is a face in Q_2, then we apply induction to Q_1 (with faces F_1 and C to be multicolored) and then to Q_2 (with multicolored C and F_2). By permuting the colors in Q_2, the two colorings coincide on C, and their union gives a required coloring of Q.

Suppose now that F_2 is in Q_1. Then we first apply induction on Q_1 with F_1, F_2 to be multicolored. The face C is not multicolored. We may assume that its coloring is 1213 or 1212. Let $x \in V(C)$ be the vertex of color 3 (if it exists). Now we color Q_2 with colors 1 and 2 (except for the vertex x) so that its coloring on C coincides with the coloring from Q_1. The colorings of Q_1 and Q_2 can now be combined to obtain a required coloring of Q.

Suppose now that every 4-cycle in Q is facial. There is a face $F = xyzw$ distinct from F_1 and F_2. Now, let Q' be obtained by identifying x and z and squeezing F. Since every 4-cycle is facial, Q' is a quadrangulation. However, we have to be certain that Q' does not contain parallel edges (since then the claim may not hold any more). By excluding parallel edges, we also make sure that each of the faces F_1 and F_2 has four distinct vertices in Q'. If Q' has parallel edges, then x and z have a common neighbor $u \notin \{y, w\}$. Since Q is planar and bipartite, y and w cannot have a common neighbor distinct from x and z. Therefore, we can squeeze F by identifying y and w instead.

By applying the induction hypothesis to Q', we get a coloring of Q' which can also be used as a required coloring of Q. □

3 Quadrangulations of the Projective Plane

Let G be a graph embedded in a surface S. If C is a cycle in G that is contractible in the surface, then C bounds a disk in S. That disk (together with its boundary C) is called the *interior* of C.

In this section we prove an extension of known results about 4-colorings of nonbipartite quadrangulations of the projective plane obtained in [You96, HRS02, Moh02, CT04]. Claim (a) of Theorem 3.1 is reproduced from [Moh02]. Part (b) extends a result that was obtained previously (with an essentially different proof) by Collins and Tysdal [CT04].

In the proof of Theorem 3.1, we will use the following easy fact. In a nonbipartite quadrangulation of the projective plane, all contractible cycles have even length and all noncontractible cycles have odd length. In particular, if we allow multiple edges, they can never form a noncontractible 2-cycle.

Theorem 3.1. *Let Q be a nonbipartite quadrangulation of the projective plane, and let k be an integer.*

(a) *If Q is k-colored, then there are at least three multicolored faces. In particular, $k \geq 4$.*
(b) *If $k = 4$, then for every type $t \in \{1, 2, 3\}$, the number of multicolored faces of type t is odd.*

Proof. Suppose that Q is not bipartite, that it is k-colored, and that it is a counterexample to either (a) or (b) with the minimum number of vertices. Let $\mathcal{F}_1$ be the set of multicolored faces. Denote by $\mathcal{F}$ the set of all faces which are not in $\mathcal{F}_1$.

We allow multiple edges. However, we shall prove that they do not appear in Q. If Q would have a pair of edges joining vertices x and y, the corresponding 2-cycle C would be contractible since all noncontractible cycles are odd. Hence, C would bound a disk D. By deleting the interior of D and identifying the two parallel edges, we would get a smaller counterexample (where we apply Theorem 2.1 to D when proving (b)), which would contradict our choice of Q.

Suppose now that Q has a facial walk $xyzw \in \mathcal{F}$ such that x and z have the same color. Clearly, $x \neq z$ since Q has no parallel edges. Then we can squeeze the face by identifying x and z. The resulting graph is a nonbipartite k-colored quadrangulation of the projective plane with the same multicolored faces, which yields a contradiction to the minimality of Q.

The conclusion of the above is that $\mathcal{F} = \emptyset$ (and that Q does not have multiple edges). Since $\mathcal{F}_1 \neq \emptyset$, the quadrangulation Q has $n \geq 4$ vertices. For quadrangulations on the projective plane, Euler's formula implies that the number of faces is $n-1 \geq 3$. This proves (a). As it only remains to prove (b), we assume henceforth that $k=4$.

Now we proceed in the same way as in the proof of Theorem 2.1. First we exclude the case when two faces of the same type share an edge. Having done that, we conclude that the dual graph of Q covers the dual graph R of the K_4-quadrangulation. Observe that R has three vertices and that any pair of them is joined by two edges forming a noncontractible 2-cycle in the projective plane. It follows that the number of faces of Q is $3d$, where d is the degree of the covering projection. If $d=1$, then $Q=K_4$, which has a unique 4-coloring, with one multicolored face of each type. This concludes the proof for $d=1$.

If $d>1$, then Q has a vertex u of degree 3 which is adjacent to a vertex v whose degree is more than 3. (This follows from Euler's formula by using standard counting arguments.) In this case, the coloring around u and v is as shown in Figure 5, and we can make a reduction shown in that figure, and finally apply the induction hypothesis. There are some minor technical details about this reduction that are worth mentioning. First, the reduction shown in Figure 5 gives rise to a loopless 4-colored graph since the added edges join vertices of distinct colors. However, there is a possible trouble if $x=u$. In that case, the vertex x is not present after the deletion of u. If this happens, we apply a similar reduction at the vertex y (see Figure 5). Let us observe that in this case $y \neq u$, since if it were, u would be contained in four quadrangular faces, contradicting the fact that its degree is 3. Lastly, by performing this reduction, four multicolored faces of types 1, 2, 2, 3 are replaced with two multicolored faces, whose types are 1 and 3, and one face which is not multicolored. Hence, the parities of the numbers of multicolored faces of specific types remain unchanged.

This completes the proof. □

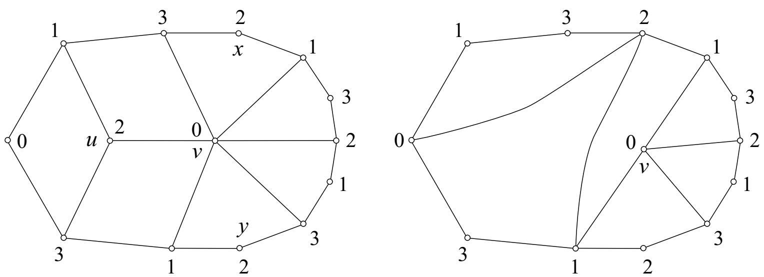


Fig. 5. Another reduction

Let Q be a quadrangulation and $\mathcal{F}$ a collection of some of the faces of Q. The pair $(Q,\mathcal{F})$ is called a *bordered quadrangulation*, and we think of it as being the 2-cell complex obtained from Q by deleting the 2-cells corresponding to the faces in $\mathcal{F}$. A 4-coloring of the bordered quadrangulation is *soft* if all faces of Q distinct from those in $\mathcal{F}$ have at most three colors. Observe that this notion is natural when passing from (bordered) quadrangulations to (bordered) triangulations in which each face is triangulated. A 4-coloring of Q can be extended to a 4-coloring of the corresponding triangulations if and only if it is soft. See Section 4.

A corollary of Theorem 3.1(b) is the following

Lemma 3.2. *Let A, B, C be distinct faces of a nonbipartite quadrangulation Q of the projective plane. If $B = xyzw$ and $xy \subseteq A \cap B$, $zw \subseteq B \cap C$, then $(Q, \{A, B, C\})$ has no soft 4-coloring.*

Proof. If c is a soft 4-coloring of $(Q, \{A, B, C\})$, then A, B, C are all multicolored by Theorem 3.1(b), and they are of diffierent types. We may assume that $c(x) = 0$, $c(y) = 1$, $c(z) = 2$, $c(w) = 3$, so B is of type 2. Then A must be of type 3. Now, color 1 is opposite to z or opposite to w in C. It follows that C cannot be of type 1, a contradiction. □

Now we shall add some specific examples.

Lemma 3.3. *Let A, B, C, D be the faces of the projective planar quadrangulation Q_0 as shown in Figure 1(a). Then the bordered quadrangulation $(Q_0, \{A, B, C, D\})$ has no soft 4-colorings. On the other hand, if $F \notin \{A, B, C, D\}$ is another face of Q_0, then $(Q_0, \{A, B, C, F\})$ admits a soft 4-coloring.*

Proof. Soft colorings of $(Q_0, \{A, B, C, F\})$ are shown in Figure 6 for different choices of F (up to symmetries). The types of the three multicolored faces are shown inside small circles.

To prove the first part of the lemma, let $A = xyzw$, where $x \in V(D)$, $B = wzts$ and $C = stuv$, where $v \in V(D)$ is opposite to x in D. Suppose now that $(Q_0, \{A, B, C, D\})$ has a soft 4-coloring c. By Theorem 3.1(b), precisely three of the faces A, B, C, D are multicolored. By Lemma 3.2, D is necessarily one of them, and we will assume that it is of type 1 and that $c(x) = 0$ and $c(v) = 1$.

Suppose first that A, B, D are multicolored. Since A is not of type 1 and colors 2, 3 have not yet been introduced, we may assume that A is of type 2, so that $c(z) = 2$. Then B is of type 3, and hence $c(t) = 0$ and $c(w) = 3$. Since B is multicolored, it follows that $c(s) = 1$, a contradiction since its neighbor v is colored 1.

The case when B, C, D are multicolored is symmetric to the above, so it remains to consider the case when A, C, D are multicolored. Again, we may assume that A is of type 2, so $c(z) = 2$. Since C is of type 3, we conclude that $\{c(s), c(u)\} = \{0, 3\}$ and $c(t) = 2$. But this is a contradiction since the neighbor z of t is also colored 2. □

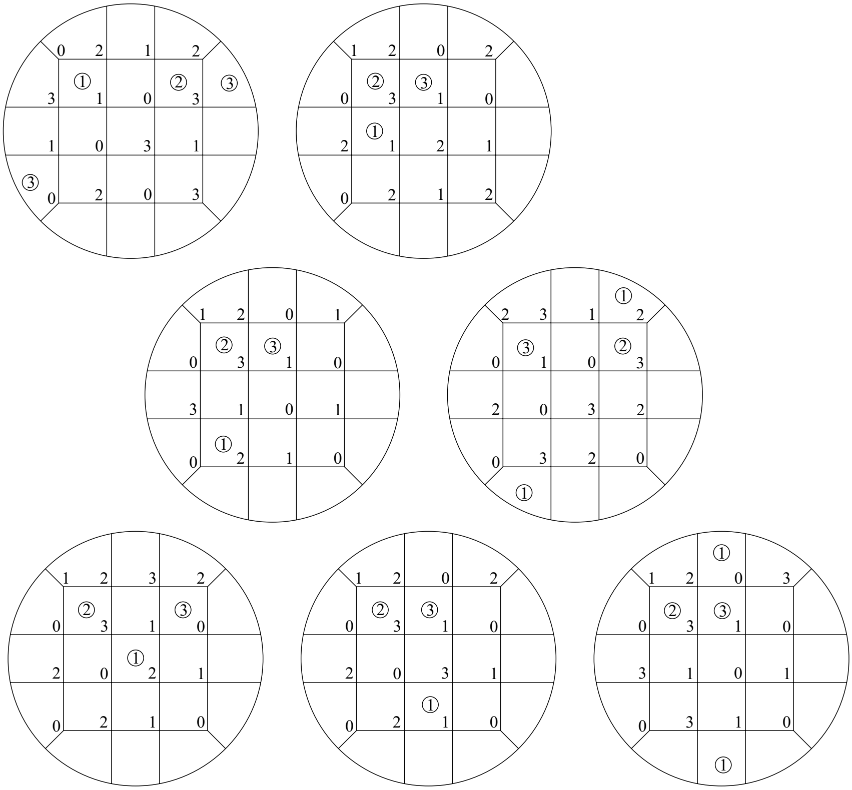


Fig. 6. Soft colorings of $(Q_0, \{A, B, C, F\})$

A bordered quadrangulation $(Q, \mathcal{F})$ is *soft-4-critical* if it does not have a soft 4-coloring but for every face F of Q, where $F \notin \mathcal{F}$, $(Q, \mathcal{F} \cup \{F\})$ has a soft 4-coloring.

Some soft-4-critical bordered quadrangulations of the form $(Q_0, \mathcal{F})$ are presented in Figure 7, where the faces in $\mathcal{F}$ are represented by circles. Criticality of the first one is a direct consequence of Lemma 3.3. For the other two, the proof is similar, and we leave details to the reader.

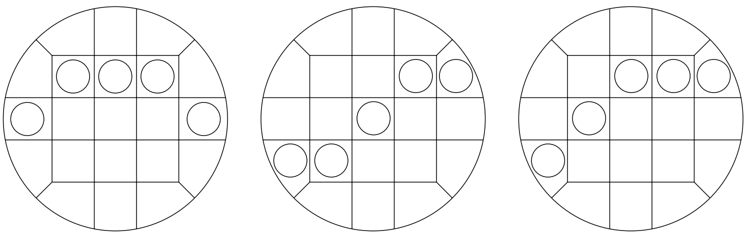


Fig. 7. Some soft-4-critical bordered quadrangulations

An unfortunate property of the above examples is that they cannot be used directly to construct soft-4-critical bordered quadrangulations of large edge-width. This is repaired in the examples given in the sequel.

Let Q_0 be the quadrangulation shown in Figure 1(a). For $k \geq 1$, let Q_k be the projective planar quadrangulation obtained from Q_0 as follows. First, subdivide the six edges e_i ($1 \leq i \leq 6$) passing "through the crosscap" (the outside edges), each with $2k$ new vertices $v_{i,j,l}$, where $1 \leq j \leq k$ and $l = 1, 2$. These vertices subdivide e_i in the respective order $v_{i,1,1}, v_{i,2,1}, \ldots, v_{i,k,1}, v_{i,k,2}, \ldots, v_{i,1,2}$. Finally, we add k cycles of lenth 12, each surrounding the "outer" 12-cycle of the 3×3 grid in Q_0. The jth cycle passes through vertices $v_{1,j,1}, \ldots, v_{6,j,1}, v_{1,j,2}, \ldots, v_{6,j,2}$. In Figure 1, quadrangulations Q_0 and Q_1 are shown.

Theorem 3.4. *Let $k \geq 1$ be an integer, and let A, B, C be the faces of the projective quadrangulation Q_k as shown in Figure 1. Then the bordered quadrangulation $(Q_k, \{A, B, C\})$ is soft-4-critical.*

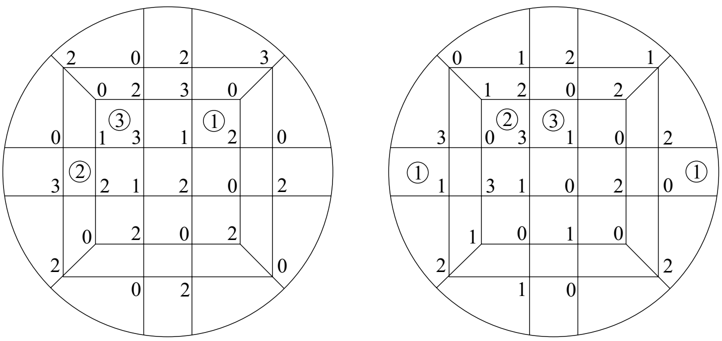


Fig. 8. Soft colorings for D_0 and D_1

Proof. Let us first observe that Q_1 is obtained from Q_0 by replacing each of the six "exterior" faces $F_1, \ldots, F_6$ by three quadrangles, i.e., they are replaced by subdividers M shown in Figure 2. Similarly, Q_k ($k \geq 2$) is obtained from Q_{k-1} by replacing six faces by subdividers M. In the sequel, we give details for the proof that Q_1 is soft-4-critical, and leave the general case to the reader.

Let $F_1, \ldots, F_6$ be the "exterior" faces of Q_0, and let $e_i = a_i b_i$ be the edge shared by F_{i-1} and F_i ($1 \leq i \leq 6$; all indices are considered modulo 6). In other words, the "outer" 12-cycle of the central 3×3-grid in Q_0 is $a_1 a_2 \ldots a_6 b_1 b_2 \ldots b_6$. In Q_1, e_i is subdivided by two vertices a'_i, b'_i, where we assume that b'_i is adjacent to a_i, and a'_i is adjacent to b_i. Here we slightly adapt the convention about the notation modulo 6 and assume that $(a_1, b'_1, a'_1, b_1) = (b_7, a_7,' b'_7, a_7)$.

Suppose that c is a 4-coloring of Q_0 with the following properties:

(a) At most one of the faces F_i, $1 \le i \le 6$, is multicolored.
(b) If $F = F_i = a_i a_{i+1} b_{i+1} b_i$ is multicolored, then either the color $c(a_i)$ appears twice in F_{i-1} and $c(b_{i+1})$ appears twice in F_{i+1}, or $c(b_i)$ appears twice in F_{i-1} and $c(a_{i+1})$ appears twice in F_{i+1}.

Now we extend c to Q_1 as follows. If neither F_i nor F_{i-1} is multicolored, we set $c(a_i') := c(a_i)$ and $c(b_i') := c(b_i)$. If F_i is multicolored, then we assume that the color $c(a_i)$ appears twice in F_{i-1} and $c(b_{i+1})$ appears twice in F_{i+1}. (The other possibility provided by (b) can be handled similarly.) Then we set $c(a_i') := c(a_i)$ and $c(b_{i+1}') := c(b_{i+1})$. This choice guarantees that the new faces in F_{i-1} and F_{i+1} will not be multicolored. For the colors of b_i' and a_{i+1}' we use one of the following three possibilities: twice the color $c(b_i)$, colors $c(a_{i+1})$ and $c(b_i)$ (respectively), or twice the color $c(a_{i+1})$. Each of these three possibilities gives rise to a different multicolored subface of F_i.

The described extension of the colorings from Figure 6 (which satisfy (a) and (b)) prove that $(Q_1, \{A, B, C, F\})$ has a soft 4-coloring for all faces $F \notin \{A, B, C, D_0, D_1, D_2\}$ (where the faces D_0, D_1, D_2 are those shown in Figure 1(b). Finally, for $F \in \{D_0, D_1\}$, the corresponding soft 4-colorings are exhibited in Figure 8, and the case of $F = D_2$ is symmetric to the case when $F = D_0$. The proof is complete. □

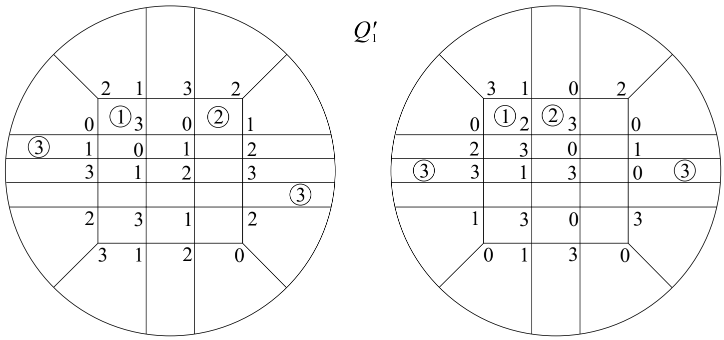


Fig. 9. Two 4-colorings of the projective quadrangulation Q_1'

Another family of quadrangulations can be derived from Q_0. If we replace each of the four faces in the "middle" by the subdivider M (see Figure 2), we obtain Q_1', which is shown in Figure 9. If we repeat the same with the new four faces forming the middle Möbius strip, we obtain Q_2'. By doing this k times all together, we get Q_k' $(k \ge 1)$.

Theorem 3.5. *Let $k \ge 1$ be an integer, and let A, B, C be the faces of the projective quadrangulation Q_k' corresponding to the faces of Q_0 shown in Figure 1. Then the bordered quadrangulation $(Q_k', \{A, B, C\})$ is soft-4-critical.*

Proof. The proof is similar to the proof of Theorem 3.4. The necessary soft 4-colorings of $(Q_1', \{A, B, C, D_0\})$ and $(Q_1', \{A, B, C, D_1\})$ are shown in Figure 9 (where D_0 and D_1 are the multicolored faces shown in the figure). The details are omitted. □

4 Eulerian Triangulations

A graph is *Eulerian* if all its vertices have even degree. It is well known that Eulerian triangulations of the plane are 3-colorable. However, Eulerian triangulations on other surfaces may have arbitrarily large chromatic number. It is easy to find examples on the projective plane whose chromatic number is equal to 3, 4, or 5, respectively, and it is easy to see that the chromatic number of an Eulerian triangulation of the projective plane cannot be more than 5. In [Moh02], a simple characterization and a polynomial time algorithm are given to decide if an Eulerian triangulation of the projective plane is 3-, 4-, or 5-colorable.

The class of graphs embedded in some surface S such that all facial walks have even length (called *locally bipartite embeddings*) is closely related to Eulerian triangulations of S. Namely, if we insert a new vertex in each of the faces of a locally bipartite embedded graph G, and join it to all vertices on the corresponding facial walk, we obtain an Eulerian triangulation $\mathcal{T}(G)$ which contains G as a subgraph. We say that $\mathcal{T}(G)$ is a *face subdivision* of G and that the set of added vertices $U = V(\mathcal{T}(G)) \setminus V(G)$ is a *color factor* of $\mathcal{T}(G)$. Since U is an independent set, $\chi(G) \le \chi(\mathcal{T}(G)) \le \chi(G) + 1$, where $\chi(\cdot)$ denotes the chromatic number of the corresponding graph.

Theorem 3.1 applied to a nonbipartite projective planar quadrangulation Q implies that the chromatic number of its face subdivision $\mathcal{T}(Q)$ is equal to 5. Theorem 3.1 also implies that $\mathcal{T}(Q)$ is not 5-critical since the removal of any two vertices of degree 4 in $\mathcal{T}(Q)$ leaves a graph which is not 4-colorable.

Eulerian triangulations of the projective plane with chromatic number 5 may have arbitrarily large face-width and they show that nonorientable surfaces behave differently than the orientable ones. Namely, Hutchinson, Richter, and Seymour [HRS02] proved that Eulerian triangulations of orientable surfaces of sufficiently large face-width are 4-colorable.

The concept of face subdivisions extends to bordered surfaces. Given a bordered quadrangulation $(Q, \mathcal{F})$, we define $\mathcal{T}(Q, \mathcal{F})$ in the same way as above, except that we do not subdivide the faces in $\mathcal{F}$. If c is a 4-coloring of $\mathcal{T}(Q, \mathcal{F})$, then its restriction to Q is a soft 4-coloring of $(Q, \mathcal{F})$. Conversely, every soft 4-coloring of $(Q, \mathcal{F})$ can be extended to a 4-coloring of $\mathcal{T}(Q, \mathcal{F})$.

Our next result shows that soft-4-criticality of bordered quadrangulations is essentially equivalent to 5-criticality of their face subdivisions.

Theorem 4.1. *Let* $(Q, \mathcal{F})$ *be a soft-*4*-critical bordered quadrangulation of the projective plane. If every (contractible)* 4*-cycle of* Q *bounds a face of* Q*, then the graph of the face subdivision* $\mathcal{T}(Q, \mathcal{F})$ *is* 5*-critical.*

Proof. Since $(Q, \mathcal{F})$ has no soft-4-colorings, $T = \mathcal{T}(Q, \mathcal{F})$ cannot be 4-colored. Thus, it suffices to see that the removal of any edge uv of T yields a 4-colorable graph.

Suppose first that u is a vertex which is not in Q, i.e., u subdivides some face $F \notin \mathcal{F}$. Soft-4-criticality of $(Q, \mathcal{F})$ implies that there is a soft 4-coloring of $(Q, \mathcal{F} \cup \{F\})$. This 4-coloring can be extended to a 4-coloring of $T - u$ since it is soft, and it can further be extended to u in $T - uv$ since u has degree 3 in $T - uv$. This proves that $T - uv$ is 4-colorable.

Suppose now that $uv \in E(Q)$. Since every contractible 4-cycle of Q bounds a face in Q, $Q - uv$ is 3-colorable, as proved by Gimbel and Thomassen [GT97] (cf. also [MS02]). Obviously, a 3-coloring of $Q - uv$ can be extended to a 4-coloring of $T - uv$. This completes the proof. □

Theorem 4.1 (together with results of Section 3) gives rise to 5-critical graphs on the projective plane. By Theorems 3.4 and 3.5, graphs $\mathcal{T}(Q_k, \{A, B, C\})$ and $\mathcal{T}(Q'_k, \{A, B, C\})$ are 5-critical for every $k \geq 1$. Theorem 4.1 and Lemma 2.3 imply that adding subdividers K and L in faces distinct from A, B, C in such a way that another quadrangulation is produced, yields new soft-4-critical bordered quadrangulations. Consequently, new 5-critical graphs are obtained as their face subdivisions. This in particular proves Theorem 1.1.

Let us observe that the edge-width of $\mathcal{T}(Q'_k, \{A, B, C\})$ can also be made arbitrarily large by using the subdividers K and L (indeed only K suffices). Both of these constructions yield the first explicit families of 5-critical graphs of arbitrarily large edge-width on a fixed surface.

References

[CT04] K. L. Collins, K. Tysdal, Dependent edges in Mycielski graphs and 4-colorings of 4-skeletons, *J. Graph Theory* 46 (2004), 285–296.

[Edw92] K. Edwards, The complexity of some graph colouring problems, *Discrete Appl. Math.* 36 (1992), 131–140.

[Fisk78] S. Fisk, The nonexistence of colorings, J. Combinatorial Theory Ser. B 24 (1978) 247–248.

[GJ79] M. R. Garey, D. S. Johnson, *Computers and Intractability: A Guide to the Theory of the NP-Completeness*, W. H. Freeman, San Francisco, 1979.

[GT97] J. Gimbel, C. Thomassen, Coloring graphs with fixed genus and girth, *Trans. Amer. Math. Soc.* 349 (1997), 4555–4564.

[HRS02] J. Hutchinson, R. B. Richter, P. Seymour, Colouring Eulerian triangulations, *J. Combin. Theory Ser. B* 84 (2002), 225–239.

[Moh93] B. Mohar, 7-critical graphs of bounded genus, *Discrete Math.* 112 (1993), 279–281.

[Moh02] B. Mohar, Coloring Eulerian triangulations of the projective plane, *Discrete Math.* 244 (2002), 339–343.

[MT01] B. Mohar, C. Thomassen, *Graphs on Surfaces*, The Johns Hopkins University Press, Baltimore and London, 2001.

[MS02] B. Mohar, P. D. Seymour, Coloring locally bipartite graphs on surfaces, *J. Combin. Theory, Ser. B* 84 (2002), 301–310.

[RSST96] N. Robertson, D. P. Sanders, P. Seymour, R. Thomas, Efficiently four-coloring planar graphs, *Proc. ACM Symp. Theory Comput.* 28 (1996), 571–575.

[Tho97] C. Thomassen, Color-critical graphs on a fixed surface, *J. Combin. Theory Ser. B* 70 (1997), 67–100.

[You96] D. A. Youngs, 4-chromatic projective graphs, *J. Graph Theory* 21 (1996), 219–227.

Crossing Number of Toroidal Graphs

János Pach[1*] and Géza Tóth[2**]

[1] Rényi Institute, Hungarian Academy of Sciences, H-1364 Budapest, POB 127, Hungary, and Department of Computer Science, City College, CUNY, USA
pach@renyi.hu

[2] Rényi Institute, Hungarian Academy of Sciences, H-1364 Budapest, POB 127, Hungary
geza@renyi.hu

Summary. It is shown that if a graph of n vertices can be drawn on the torus without edge crossings and the maximum degree of its vertices is at most d, then its planar crossing number cannot exceed cdn, where c is a constant. This bound, conjectured by Brass, cannot be improved, apart from the value of the constant. We strengthen and generalize this result to the case when the graph has a crossing-free drawing on an orientable surface of higher genus and there is no restriction on the degrees of the vertices.

AMS Subject Classification. 05C10, 57M15.
Keywords. Crossing number, genus, Menger theorem.

1 Introduction

Let S_g be the compact orientable surface with no boundary, of genus g. Given a simple graph G, a *drawing* of G on S_g is a representation of G such that the vertices of G are represented by points of S_g and the edges are represented by simple (i.e., non-selfintersecting) continuous arcs in S_g, connecting the corresponding point pairs and not passing through any other vertex. The *crossing number* of G on S_g, $\mathrm{cr}_g(G)$, is defined as the minimum number of edge crossings over all drawings of G in S_g. For $\mathrm{cr}_0(G)$, the "usual" planar crossing number, we simply write $\mathrm{cr}(G)$.

Let G be a graph of n vertices and e edges, and suppose that it can be drawn on the torus without crossing, that is, G satisfies $\mathrm{cr}_1(G) = 0$. How large can $\mathrm{cr}(G)$ be? Clearly, we have $\mathrm{cr}(G) < \binom{e}{2}$, and this order of magnitude can be attained, as shown by the following example. Take five vertices and

* Supported by NSF grant CCF-05-14079 and grants from NSA, PSC-CUNY, Hungarian Research Foundation, and BSF.
** Supported by OTKA-T-038397 and OTKA-T-046246.

connect any pair of them by $\frac{e}{20}$ vertex-disjoint paths of lengths two. In any drawing of this graph in the plane, every subdivision of K_5 gives rise to a crossing. Therefore, the number of crossings must be at least $\frac{e^2}{400}$.

Peter Brass suggested that this estimate can be substantially improved if we impose an upper bound on the degree of the vertices. More precisely, we have

Theorem 1. *Let G be a graph of n vertices with maximum degree d, and suppose that G has a crossing-free drawing on the torus. Then we have $\mathrm{cr}(G) \leq cdn$, where c is a constant.*

For $d \geq 3$, the bound in Theorem 1 cannot be improved, apart from the value of the constant c. Consider the following example. Let $d \geq 4$, $G = C_k \times C_k$, where $k = \sqrt{n/d}$ is a large integer and C_k denotes a cycle of length k. Obviously, this graph can be drawn on the torus without crossings. On the other hand, by a result of Salazar and Ugalde [SU04], its planar crossing number is larger than $(\frac{4}{5} - \varepsilon)k^2$, for any $\varepsilon > 0$, provided that k is large enough. Substitute every edge e of G by $\lfloor \frac{d}{4} \rfloor$ new vertices, each connected to both endpoints of e. The resulting graph G' has at most n vertices, each of degree at most d. It can be drawn on the torus with no crossing, and its planar crossing number is at least

$$\left(\frac{4}{5} - \varepsilon\right) k^2 \times \left\lfloor \frac{d}{4} \right\rfloor^2 > \frac{1}{100} nd.$$

To see this, it is enough to observe that there is an optimal drawing of G' in the plane with the property that any two paths of length two connecting the same pair of vertices cross precisely the same edges. The same construction can be slightly modified to show that $\mathrm{cr}(G)$ can also grow linearly in n if the maximum degree d is equal to three.

Theorem 1 can be generalized as follows.

Theorem 2. *Let G be a graph of n vertices of maximum degree d that has a crossing-free drawing on S_g, the orientable surface of genus g. Then we have $\mathrm{cr}(G) \leq c_{d,g} n$, where $c_{d,g}$ is a constant depending on d and g.*

We can drop the condition on the maximum degree and obtain an even more general statement.

Theorem 3. *Let G be a graph of n vertices with degrees $d_1, d_2, \ldots, d_n$, and suppose that G has a crossing-free drawing on S_g. Then we have*

$$\mathrm{cr}(G) \leq c_g \sum_{i=1}^{n} d_i^2,$$

where c_g is a constant depending on g.

To simplify the presentation and to emphasize the main idea of the proof, in Section 2 first we settle the simplest (planar) case (Theorem 1). In Section 3, we reduce Theorem 3 to a similar upper bound on the crossing number of G in S_{g-1} (Theorem 3.1). This latter result is established in Section 4.

2 The Simplest Case: Proof of Theorem 1

We can assume that $d \geq 3$. It is sufficient to prove that $\mathrm{cr}(G) \leq cd(n-1)$ holds for any *two-connected* graph G satisfying the conditions. Indeed, if G is disconnected or has a cut vertex, then it can be obtained as the union of two graphs G_1 and G_2 with n_1 and n_2 vertices that have at most one vertex in common, so that we have $n_1 + n_2 = n$ or $n+1$. Arguing for G_1 and G_2 separately, we obtain by induction that

$$\mathrm{cr}(G) = \mathrm{cr}(G_1) + \mathrm{cr}(G_2) \leq cd(n_1 - 1) + cd(n_2 - 1) \leq cd(n-1),$$

as required.

Let G be a two-connected graph with maximum degree d and $\mathrm{cr}_1(G) = 0$. Fix a crossing-free drawing of G on the torus. We can assume that the boundary of each face is connected. Indeed, if one of the faces contains a cycle not contractible within the face, then cutting the torus along this cycle we do not damage any edge of G. Therefore, G is a planar graph and there is nothing to prove.

If our drawing is not a triangulation, then by adding $O(n)$ extra vertices and edges we can turn it into one so that the maximum degree of the vertices increases by at most a factor of three. We have to apply the following easy observation.

Lemma 2.1. *Let G be a two-connected graph with n vertices of degree at most d ($d \geq 3$). Suppose that G has a crossing-free drawing on the orientable surface of genus g such that the boundary of each face is connected. Any such drawing can be extended to a triangulation of the surface with at most $19n + 36(g-1)$ vertices of maximum degree at most $3d$.*

Proof. First consider a cycle $f = x_1 x_2 \ldots x_{n(f)}$ bounding a single face in the drawing of G. Note that some vertices $x_i \in V(G)$ and even some edges may appear along this cycle several times. Take a simple closed curve $\gamma_0 = p_1 p_2 \ldots p_{n(f)}$ inside the face, running very close to f and passing through the (new) points p_i in this cyclic order. In the ring between f and γ_0, connect each vertex x_i to p_i and p_{i+1} (where $p_{n(f)+1} := p_1$).

Divide γ_0 into $m_0 := \lceil \frac{n(f)}{d-1} \rceil$ connected pieces, each consisting of at most d vertices, such that the last vertex of each piece π_i is the first vertex of π_{i+1}, where $1 \leq i \leq m_0$ and $\pi_{m_0+1} := \pi_1$. Place a simple closed curve $\gamma_1 = q_1 q_2 \ldots q_{m_0}$ in the interior of γ_0. In the ring between γ_0 and γ_1, connect each q_i to all points in π_i. (If $m_0 = 1$ or 2, then γ_1 degenerates into a point or a single edge.) If γ_1 has more than three vertices, repeat the same procedure for γ_1 in the place of γ_0, and continue as long as the interior of the face is not completely triangulated. We added

$$n(f) + m_0 + m_1 + \ldots < n(f) + n(f) + \frac{n(f)}{2} + \frac{n(f)}{4} + \ldots < 3n(f)$$

new vertices, and their maximum degree is at most $d+4$. The degree of every original vertex of f increased by at most twice the number of times it appeared in f.

If we triangulate every face of G in the above manner, the resulting drawing G' defines a triangulation of the surface with fewer than $n + \sum_f 3n(f) \leq n + 6|E(G)|$ vertices, each of degree at most $d' := 3d$. By Euler's formula, we have $n + 6|E(G)| \leq n + 18(n - 2 + 2g)$, as required. □

In the sequel, slightly abusing the notation, we write G for the triangulation G' and d for its maximum degree d'.

If G has no *noncontractible* cycle, i.e., no cycle represented on the torus by a closed curve not contractible to a point, then we are done, because G is a planar drawing so that $\mathrm{cr}(G) = 0$. Otherwise, choose a noncontractible cycle C with the minimum number of vertices, fix an orientation of C, and let $k := |V(C)|$. Let E_l (and E_r) denote the set of edges not belonging to C that are incident to at least one vertex of C and in a small neighborhood of this vertex lie on the left-hand side (respectively right-hand side) of C. Note that the sets E_l and E_r are disjoint, but this fact is not necessary for the proof.

Replace C by two copies, C_r and C_l, lying on its right-hand side and left-hand side. Connect each edge of E_r (respectively E_l) to the corresponding vertex of C_r (respectively C_l). Cut the torus along C, and attach a disk to each side of the cut.

The resulting spherical (planar) drawing G_1 represents a graph, slightly different from G. To transform it into a drawing of G, we have to remove C_l and (re)connect the edges of E_l to the corresponding vertices of C_r. In what follows, we describe how to do this without creating too many crossings.

Let $\hat{G}_1$ denote the *dual* graph of G_1, that is, place a vertex of $\hat{G}_1$ in each face of G_1, and for any $e \in E(G_1)$ connect the two vertices assigned to the faces meeting at e by an edge $\hat{e} \in E(\hat{G}_1)$. Let r and l denote the vertices of $\hat{G}_1$ lying in the faces bounded by C_r and C_l.

Lemma 2.2. *In $\hat{G}_1$, there are k vertex-disjoint paths between the vertices r and l.*

Proof. By Menger's theorem, the maximum number p of (internally) vertex-disjoint paths connecting r and l in $\hat{G}_1$ is equal to the minimum number of vertices whose deletion separates r from l. Choose p such separating vertices, and denote the corresponding triangular faces of G by $f_1, \ldots, f_p$. The interior of the union of these faces must contain a noncontractible closed curve that

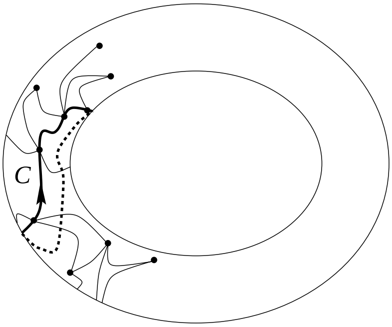


Fig. 1. C is the shortest noncontractible cycle

does not pass through any vertex of G. Let δ be such a curve whose number of intersection points with the edges of G is minimum. Choose an orientation of δ. Let $e_1, \ldots, e_q$ denote the circular sequence of edges of G intersected by δ. By the minimality of δ, we have $q \leq p$, because the interior of each triangle f_i contains at most one maximal connected piece of δ. Let v_i be the right endpoint of e_i with respect to the orientation of δ. Notice that v_i is adjacent to or identical with v_{i+1}, for every $1 \leq i \leq q$ (where $v_{q+1} := v_1$). Therefore, the circular sequence of vertices $v_1, \ldots, v_q$ induces a cycle in G that can be continuously deformed to δ. Thus, we have a noncontractible cycle of length $q \leq p$ in G, which implies that k, the length of the shortest such cycle, is at most p, as required. □

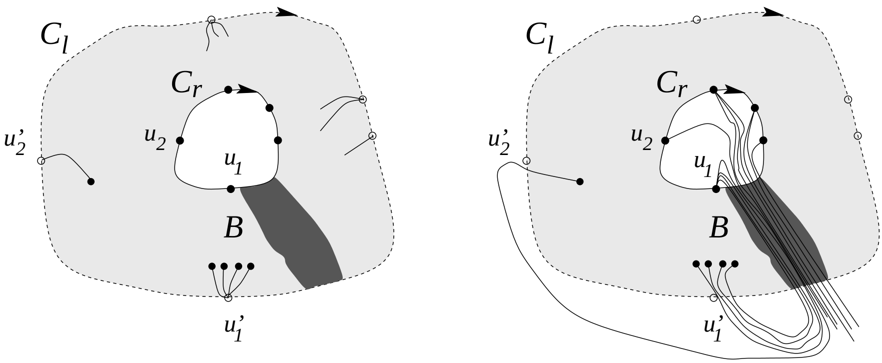


Fig. 2. Pulling the edges in E_l through the corridor B

By Lemma 2.1, the graph $\hat{G}$ has at most $2|V(G)| \leq 38n$ vertices. According to Lemma 2.2, there is a path connecting r and l in $\hat{G}$ with fewer than $\frac{38n}{k}$ internal vertices. The corresponding faces of G_1 form a "corridor" B between

C_r and C_l. Delete now the vertices of C_l from G_1. Pull every edge in E_l through B, and connect each of them to the corresponding vertex of C_r. See Figures 1 and 2. Notice that during this procedure one can avoid creating any crossing between edges belonging to E_l.

We give an upper bound on the number of crossings in the resulting planar drawing of G. Using that $|C| = k$ and $|E_l| \leq dk$, we can conclude that by pulling each edge through the corridor B, we create at most $\frac{38n}{k}$ crossings per edge. Thus, the total number of crossings cannot exceed $dk \cdot \frac{38n}{k} = 38dn$, which completes the proof of Theorem 1. □

3 Reducing Theorem 3 to Theorem 3.1

Given a graph G, let $n(G)$ and $\sigma(G)$ denote the number of vertices of G and the sum of the squares of their degrees.

Theorem 3 provides an upper bound for the crossing number of a graph G that can be drawn on S_g without crossing. Next we show that this bound can be deduced by repeated application of the following result. In each step, we reduce the genus of the surface by one.

Theorem 3.1. *Let G be a two-connected graph with $\mathrm{cr}_g(G) = 0$. Then we have $\mathrm{cr}_{g-1}(G) \leq c_g^* \sigma(G)$, for some constant $c_g^* \geq 1$.*

Proof of Theorem 3 using Theorem 3.1. As in the proof of Theorem 1, we can assume that G is two-connected. Consider a crossing-free drawing of $G_0 := G$ on S_g. According to Theorem 3.1, G_0 can be drawn on S_{g-1} with at most $c\sigma(G)$ crossings. Place a new vertex at each crossing, and apply Theorem 3.1 to the resulting graph G_1. Proceeding like this, we obtain a series of graphs $G_2, G_3, \ldots, G_g$, drawn on $S_{g-2}, S_{g-3}, \ldots, S_0$, respectively, with no crossing.

We claim that for any i, $0 \leq i \leq g$,

$$\sigma(G_i) \leq (17)^i \left(\prod_{g-i<j\leq g} c_j^* \right) \sigma(G)$$

holds. This is obviously true for $i = 0$. Let $0 < i \leq g$, and assume that the claim has already been verified for $i-1$. Notice that, apart from the original vertices of G_{i-1}, every other vertex of G_i has degree four. Thus, applying Theorem 3.1 to the graph G_{i-1} that had a crossing-free drawing on S_{g-i+1}, we obtain

$$\sigma(G_i) \leq \sigma(G_{i-1}) + 16\mathrm{cr}_{g-i}(G_{i-1}) \leq \sigma(G_{i-1}) + 16c_{g-i+1}^* \sigma(G_{i-1})$$

$$\leq (1 + 16c_{g-i+1}^*)(17)^{i-1} \left(\prod_{g-i+1<j\leq g} c_j^* \right) \sigma(G) \leq (17)^i \left(\prod_{g-i<j\leq g} c_j^* \right) \sigma(G),$$

which proves the claim.

It follows from the construction that G_g is a planar graph, and we have

$$n(G_g) - n(G) < \sigma(G_g) \leq 17^g \left(\prod_{j=1}^{g} c_j^* \right) \sigma(G).$$

Replacing the $n(G_g) - n(G)$ "new" vertices of G_g by proper crossings, we obtain a drawing of G in the plane with at most $17^g \left(\prod_{j=1}^{g} c_j^* \right) \sigma(G)$ crossings. This completes the proof of Theorem 3. □

4 Reducing the Genus by One: Proof of Theorem 3.1

It remains to prove Theorem 3.1.

All noncrossing closed curves C on S_g belong to one of the following three categories:

1. C is *contractible* (to a point);
2. C is *noncontractible* and *twosided*, i.e., it separates S_g into two connected components;
3. C is *noncontractible* and *onesided*.

Let us cut the surface S_g along C, and attach a disk along each side of the cut. If C is contractible, we obtain two surfaces: one homeomorphic to S_g and the other homeomorphic to the sphere S_0. If C is noncontractible and twosided, then we obtain two surfaces homeomorphic to S_a and S_b, for some $a, b > 0$ with $a + b = g$. Finally, if C is noncontractible and onesided, then we get only *one* surface, S_{g-1} [MT01].

First we need an auxiliary statement, interesting on its own right.

Theorem 4.1. *Let G be a graph with a crossing-free drawing on S_g. If G has no noncontractible onesided cycle, then G is a planar graph.*

Proof. We follow the approach of Cairns and Nikolayevsky [CN00], developed to handle a similar problem on generalized thrackles. Let S be a very small closed neighborhood of the union of all edges of the drawing of G on S_g. Then S is a compact connected surface whose boundary consists of a finite number of closed curves. Attaching a disk to each of these closed curves, we obtain a surface S' with no boundary. We show that S' is a sphere. To verify this claim, consider two closed curves, α' and β', on S'. They can be continuously deformed into closed walks, α_1 and β_1, along the edges of G. Let α and β be the corresponding closed walks along the edges of G in the original drawing on S_g. By the assumption, α divides S_g into two parts, therefore, β crosses α an even number of times. Since the original drawing of G on S_g was crossing-free, every crossing between α and β occurs at a vertex of G. Using the fact that in the new drawing of G on S', the cyclic order of the edges incident to a vertex is the same as the cyclic order of the corresponding edges in the original drawing, we can conclude that α_1 and β_1 cross an even number of times. It is not hard to argue that then the same was true for α' and β'. Thus, S' is a surface with no boundary in which any two closed curves cross an even number of times. This implies that S' is a sphere. Consequently, we have a crossing-free drawing of G on the sphere, that is, G is a planar graph. □

Proof of Theorem 3.1. As in the previous section, let $\sigma(G)$ denote the the sum of the squared degrees of the vertices of G. A *grid* of size $k \times k$ is the cross product $P_k \times P_k$ of two paths of length k. The vertices of $P_k \times P_k$ with degrees less than four are said to form the *boundary* of the grid. The proof of Theorem 3.1 is based on the same idea as that of Theorem 1, but some important details have to be modified.

Suppose that G is a two-connected graph of n vertices, drawn on S_g without crossing. We can also assume that G has no crossing-free drawing on S_{g-1}, otherwise Theorem 3.1, is trivially true. In particular, it follows that every face of the drawing of G on S_g has a connected boundary.

Replace each vertex v of degree $d(v) > 4$ by a grid of size $d(v) \times d(v)$ and connect the edges incident to v to distinct vertices on the boundary of the grid, preserving their cyclic order. The resulting crossing-free drawing of G' has at most $\sigma(G)$ vertices, each of degree at most four. Every face has a connected boundary, so that we can apply Lemma 2.1 to turn G' into a triangulation G'' with at most $19\sigma(G) + 36(g-1)$ vertices, each of degree at most twelve. Restricting G' and G'' to any grid substituting for a vertex in G, the only difference between them is that each quadrilateral face in G' is subdivided by one of its diagonals into two triangles in G''. Color all edges along the boundaries of the grids *blue*, and all other grid and diagonal edges of G'' that lie in the interior of some grid *red*.

If G'' has no noncontractible onesided cycle, then we are done by Theorem 4.1. Otherwise, pick such a cycle C with the smallest number k of vertices. Without increasing its length too much, we can replace all red edges of C by blue edges. Indeed, the first vertex and the last vertex of any maximal red path in C must belong to the boundary of the same grid. Replace each such path by the shortest blue path connecting its first and last vertices along the boundary of the grid containing them. The resulting cycle C' is noncontractible, onesided, and its length is at most $2k$. It has no red edges, and we can assume without loss of generality that it does not intersect itself. Fix an orientation of C'.

Let E_l (and E_r) denote the set of edges not belonging to C' that are incident to at least one vertex of C' and in a small neighborhood of this vertex lie on the left-hand side (respectively right-hand side) of C'.

Replace C' by two copies, C_r and C_l, lying on its right-hand side and left-hand side. Connect each edge of E_r and E_l) to the corresponding vertex of C_r and C_l. Cut S_g along C, and attach a disk to each side of the cut. The resulting surface is S_{g-1}, and it contains a crossing-free drawing G_1 of a graph slightly different from G''. To obtain a drawing of G'' from G_1, we have to remove C_l and (re)connect the edges of E_l to the corresponding vertices of C_r without creating too many crossings.

Let $\hat{G_1}$ be the *dual* drawing of G_1 on S_{g-1}. Let r (respectively l) be the vertex of $\hat{G_1}$ lying in the face bounded by C_r (respectively C_l). Color *blue* each vertex of $\hat{G_1}$ that corresponds to a face lying inside a grid in G''.

Repeating the proof of Lemma 2.2, we obtain

Lemma 4.2. *In $\hat{G_1}$, there are k vertex-disjoint paths between the vertices r and l.* □

The number of cells in G_1 is equal to the number of cells in G'' plus 2. Therefore, by Euler's formula, $\hat{G_1}$ has at most

$$2|V(G'')|+4(g-1)+2 \leq 2\,(19\sigma(G)+36(g-1))+4(g-1)+2 < 40(\sigma(G)+2g)$$

vertices. Thus, by Lemma 4.2, there is a path $P(rl)$ between r and l, of length at most $40(\sigma(G)+2g)/k$. Replacing all blue vertices of $P(rl)$ by others, we obtain a new path $P'(rl)$, not much longer than $P(rl)$. First observe that r and l, the two endpoints of $P(rl)$, are not blue. Let $uv_1v_2\ldots v_jv$ be an interval along P such that all v_i's are blue $(1 \leq i \leq j)$, but u and v are not. Then the faces corresponding to u and v must be adjacent to the boundary of some grid in G_1. These two faces are connected by two chains of faces following the outer boundary of the grid. Replace $v_1, v_2, \ldots, v_j$ by the sequence of vertices corresponding to the shorter of these two chains. Since the degree of every vertex in G_1 is at most twelve, the length of this chain is at most $12j$. Repeating this procedure for each maximal blue interval of $P(rl)$, we obtain a new path $P'(rl)$, whose length is at most $480(\sigma(G)+2g)/k$.

The corresponding faces of G_1 form a "corridor" B between C_r and C_l. Now delete r, l, and the vertices of C_l. In the same way as in the proof of Theorem 1, "pull" all edges of E_l through B, and connect them to the corresponding vertices of C_r. This step can be carried out without creating any crossing between the edges in E_l.

Now we count the number of crossings in the resulting drawing. Since $|C'| \leq 2k$, $|E_l| \leq 20k$. Pulling them through the corridor B, we create at most $480(\sigma(G)+2g)/k$ crossings per edge, that is, altogether at most $X := 9600(\sigma(G)+2g)$ crossings.

Deleting the extra vertices and edges from G_1 and collapsing each grid into a vertex, we obtain a drawing of G on S_{g-1}, in which the number of crossings cannot exceed X. This concludes the proof of Theorem 3.1. □

References

[CN00] G. Cairns and Y. Nikolayevsky, Bounds for generalized thrackles, *Discrete Comput. Geom.* **23** (2000), 191–206.

[MT01] B. Mohar and C. Thomassen: *Graphs on surfaces*, Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 2001.

[SU04] G. Salazar and E. Ugalde: An improved bound for the crossing number of $C_m \times C_n$: a self-contained proof using mostly combinatorial arguments, *Graphs Combin.* **20** (2004), 247–253.

Regular Maps on a Given Surface: A Survey

Jozef Širáň*

Department of Mathematics, SvF, Slovak University of Technology, 81368 Bratislava, Slovak Republic,
and Department of Mathematics, University of Auckland, Private Bag 92019, Auckland, New Zealand
`siran@math.sk`

Summary. Regular maps are cellular decompositions of closed surfaces with the highest 'level of symmetry', meaning that the automorphism group of the map acts regularly on flags. We survey the state-of-the-art of the problem of classification of regular maps on a given surface and outline directions of future research in this area.

AMS Subject Classification. Primary: 05C10, 05C25, Secondary: 20F65, 57M60.
Keywords. Graph, surface, embedding, map, automorphism, regular map.

1 What Is a Map?

This question was the title of a paper by W. T. Tutte [Tut73] that appeared after more than one hundred years of fruitful research into maps on surfaces. In fact, origins of what is now known as theory of maps go back to ancient Greece. Mathematics of that period was dominated by the beauty of geometry, including fascination about polyhedra with certain regularity properties. The most prominent examples are the ubiquitous five Platonic solids. The reader may think of these as of our first examples of maps. Each of the five polyhedra has vertices, edges, and faces, and may thus be viewed as a "drawing" of a graph on the sphere. In the case of the dodecahedron, for instance, it would be a spherical "drawing" containing 20 vertices, 30 edges, and 12 pentagonal faces.

Most spatial models of the dodecahedron would have all edges of the same length and all faces bounded by congruent, regular pentagons. These geometric features are, however, not of concern in the theory of maps. From now on we will be interested only in the way how vertices, edges, and faces interact, regardless of the geometric shape of edges, faces, and their boundaries. In

* This research was sponsored in part by the U.S.-Slovak Science and Technology Joint Fund, Project Number 020/2001. Partial support from the RSNZ Marsden Fund is also gratefully acknowledged.

particular, we will not require that edges be straight-line segments, faces be flat, and their boundaries be regular polygons. Only the cell structure of the resulting object will matter.

In the course of our exposition we will assume that the reader is, at least at an elementary level, familiar with fundamentals of topology and with the concept of a compact, orientable or nonorientable, 2-dimensional surface, or, shortly, a *surface*. Also, we will assume familiarity with basic notions of graph theory and, later, group theory as well.

Thus, what is a map? Intuitively, it is "a drawing of a graph on a surface". To make this precise, let us regard graphs as topological 1-dimensional complexes. Then, an *embedding* of a graph Γ on a surface $\mathcal{S}$ is a continuous one-to-one mapping $f: \Gamma \to \mathcal{S}$. One usually identifies the image $f(\Gamma)$ on the surface $\mathcal{S}$ with the graph Γ itself. Connected components of $\mathcal{S} \setminus f(\Gamma)$ are called *faces*. The embedding f is *cellular* if every face is homeomorphic to an open disc. Note that cellularity implies connectivity of the embedded graph. Also, in the cellular case, the boundary of each face is formed by a closed walk in the embedded graph. Finally, any cellular embedding of a graph will be called a *map*.

How can one actually describe maps? In other words, what do we have to specify in order to uniquely determine an embedding of a connected graph? The answer is simple if we restrict ourselves to orientable surfaces. If a graph Γ is cellularly embedded on an orientable surface, a choice of a preferred orientation of the surface induces, at each vertex, a cyclic permutation of edges leaving that vertex. It is customary to represent an edge leaving a vertex by an arrow on the edge pointing out of the vertex – or, to think of an edge with direction, although our graphs are undirected. This way, a map on an oriented surface induces a permutation of edges with directions, such that the (entries in) cycles of the permutation exactly correspond to (edges directed out of) vertices. Such a permutation is called a *rotation*. Conversely, each rotation of a connected graph gives rise to a unique map on an oriented surface. To see this, imagine that each edge is a centre of a narrow band. A cycle of a rotation then determines the cyclic order of the bands leaving a disc neighbourhood of a vertex. As the result we obtain a *band complex* in which the graph is embedded. It just remains to fill the "holes" by cells to obtain the embedding. The situation is more complicated in the nonorientable case because of lack of global orientation. In addition to a rotation, one would have to specify the way the "local" orientations given by the cycles of the rotation interact. In terms of the band complex, this would tell us which of the bands have to be "twisted". Rather than going into further details that can be found in the monograph by J. L. Gross and T. W. Tucker [GrT87], we pass on to a different approach.

The description using rotation focuses on the graph in the first place; the embedding is then constructed with the help of a band complex. A map, however, can also be viewed as a cellular decomposition of a surface. This suggests looking for a description that would capture cellularity right at the outset. Consider a map on a surface $\mathcal{S}$, formed by an embedded graph Γ, and suppose that Γ is "drawn" on $\mathcal{S}$ in thick lines. Pick a point in the interior of each face and call it the *centre* of the face. In each face, join the centre by dashed and thin line segments with every vertex and the midpoint of every edge, respectively, on the boundary of the face. The resulting refined structure is the *barycentric subdivision* of the map. Cells of the subdivision are topological triangles called *flags*. Each flag has a thick, a thin, and a dashed side (as in Fig. 1 of section 3, ignoring labels). To describe the map in terms of flags and their interaction, let $\mathcal{F}$ be the set of flags of the map. We introduce three permutations X, Y, and Z on $\mathcal{F}$ as follows. For each flag $b \in \mathcal{F}$ and for each $W \in \{X, Y, Z\}$, the flag $W(b)$ is the unique flag different from b that shares with b the thin, the thick, or the dashed side, depending on whether W is equal to X, Y, or Z. Obviously, X, Y, Z are involutions with no fixed points, and it is easy to see that X commutes with Y. Moreover, cellularity implies that the group $\langle X, Y, Z\rangle$ is a transitive permutation group on $\mathcal{F}$. Conversely, any such permutation group gives rise to a map; we will discuss the construction in detail in an important special case in section 3. In this place we just add two remarks. First, the supporting surface of the map is orientable if and only if the subgroup generated by the two products YZ and ZX has index two in $\langle X, Y, Z\rangle$. Second, the above discussion also applies to infinite maps which we will encounter in a few places in section 2.

Now, after all, what is a map? We have seen a topological definition and a combinatorial description. The answer suggested by Tutte's paper [Tut73] is purely algebraic and follows the third way we have just outlined: *A map is a transitive permutation group generated by three fixed-point-free involutions, two of which commute.*

2 What Is a Regular Map?

If an article with this title had been written, perhaps it, too, would have found its answer in permutation groups. This will transpire after we explain the basics. Before we begin, we would like to note that the concept of regularity has a number of meanings in mathematical disciplines. Here, regularity will mean "highest level of symmetry".

How would one define a "symmetry" of a map? Taking, say, the dodecahedron again, there is a large number of "symmetries" – rotations and reflections – that preserve the solid. All of them carry vertices, edges, and faces to the corresponding objects; in particular, they carry flags onto flags. This tells us what we should be looking for in general. Let M be a map with flag set $\mathcal{F}$. We define an *automorphism* of M to be a permutation of $\mathcal{F}$ that maps pairs of flags sharing a thick (dashed, thin) side to pairs sharing the same type of side. It is easy to see that such a permutation behaves as expected: it preserves the cell structure of the map and also induces an automorphism of the embedded graph. The collection of all automorphisms of M forms, under composition of mappings, the *automorphism group* of M, denoted $Aut(M)$. The important observation to make is that the group $Aut(M)$ acts *freely* on $\mathcal{F}$, that is, for any two flags $b, b' \in \mathcal{F}$ there exists at most one $\alpha \in Aut(M)$ such that $\alpha(b) = b'$.

What is now the largest "level of symmetry" a map can have? The most one can expect is that for any ordered pair of flags there is *exactly one* automorphism that maps the first flag onto the second. Such actions of groups are known as regular actions. We therefore define a map M to be *regular* if the group $Aut(M)$ acts regularly on the flag set of M. Maps with the theoretically largest "level of symmetry" are thus the regular maps.

Another, and equivalent, definition of regularity of maps makes no prior reference to automorphisms at all. Let M be a map and let $\langle X, Y, Z \rangle$ is the corresponding transitive permutation group representing M. Then, M is regular if the group $\langle X, Y, Z \rangle$ acts freely (and hence regularly) on the flag set of M. Equivalence of the two definitions follows from known facts about general group actions and we refer to R. P. Bryant and D. Singerman [BrS85] for details. This gives an immediate one-to-one correspondence between regular maps and groups generated by three involutions, two of which commute. It also opens up numerous connections to other branches of mathematics such as hyperbolic geometry, theory of Riemann surfaces, and Galois theory, as we shall see in the next section.

The two ways of introducing regular maps arrive at the goal from two opposite directions. In the first scenario, the automorphism group of a map acts freely on the set of flags. The way to make the automorphism group the *largest* possible is to require that it act transitively on the set of flags. On the other hand, the map can be identified with a certain transitive permutation group on a set, the set of flags of the map. The way to make such a group the *smallest* possible is to stipulate that it act freely on the flag set. Both approaches meet in the concept of a regular map.

If the supporting surface of a map is orientable, one may introduce a weaker concept of regularity by focusing on orientation preserving automorphisms only. In such a case we say that the map is *orientably regular* if its group of orientation preserving automorphisms acts regularly on mutually incident vertex-edge pairs, or, equivalently, on edges with a preassigned direction. A map that is orientably regular but not regular is called *chiral*.

The algebraic viewpoint of regular maps evokes the feeling that one can actually forget about surfaces and topology. This is, to some extent, the case. Such a standpoint, however, would not be productive in general, since it would cut off a considerable supply of combinatorial and topological ideas that have contributed to the theory of regular maps in the past.

Perhaps the most famous examples of regular maps are the five Platonic polyhedra that permeate our exposition. It is the wealth of non-spherical regular maps, however, that give this topic fascinating dimensions. Such maps were considered by medieval astronomers in their attempts to explain the planetary system. Prominent examples are the stellated polyhedra that appeared in the work of J. Kepler [Kep19] as early as in 1619. More than two centuries later, maps and regular maps resurfaced in two independent and, at the beginning, unrelated streams of research. The first was driven mainly by the appearance of the four colour problem and the map colouring problem of P. J. Heawood [Hea90]. In this connection, L. Heffter [Hef98] discovered orientably regular embeddings of complete graphs of prime order. Approximately at the same time, certain three-valent regular maps on a surface of genus three were studied by F. Klein [Kle79] and W. Dyck [Dyc80] in a completely different connection – constructions of multiply periodic (or, automorphic) complex functions. As a way of a geometric representation of groups, regular maps also appeared in the monograph by W. Burnside [Bur11]. The term regular map, however, was first introduced as late as in 1927 by H. R. Brahana in [Bra27], which appears to be the first systematic treatment of the topic.

Development in the classification of regular maps on a given surface in the 20th century will be overviewed in section 4. Here we only note that foundations of modern theory of orientably regular and regular maps have been laid by G. A. Jones and D. Singerman [JoS78], and by R. P. Bryant and D. Singerman in [BrS85]. The importance of the two papers, however, and even more so of the two successive articles by G. A. Jones and D. Singerman [JoS96] and by G. A. Jones [Jon97], also lies in pointing out the fascinating connections between the theory of maps, theory of groups, geometry of surfaces, Riemann surfaces, and Galois theory, which we will briefly summarise in section 3.

Classification of regular, orientably regular, and chiral maps on a given surface thus appears to be an important problem. Besides natural significance in the theory of maps, progress towards a solution of the problem would advance knowledge and find applications in the disciplines mentioned above. The aim of this paper is to survey results that have been achieved in this direction. Following a more detailed presentation of the algebraic background, in section 4 we focus on regular maps on surfaces of relatively small genus. In section 5 we give a brief account on the classification of regular maps on surfaces of Euler characteristic $\chi = -p$ where p is a prime. We also outline the newest development regarding regular maps on surfaces with $\chi = -2p$, $-3p$, and p^2, including a discussion on maps of Zassenhaus type. In the final section 6 we mention possible generalisations to regular hypermaps.

Let us note for completeness that the problem of classification of regular and orientably regular maps has been approached from other angles as well, such as classification by the automorphism groups, by the (fixed) underlying graphs, and by families of graphs. These are out of the scope of our article and would, in fact, deserve a separate survey paper. We therefore conclude with a short selection of influential results in these areas. An enumeration of orientably regular maps with automorphism groups isomorphic to 2-dimensional projective special linear groups can be extracted from results of S. H. Sah [Sah69]. An abstract characterization of graphs underlying regular and orientably regular maps was given by A. Gardiner, R. Nedela, M. Škoviera and the author in [GNS99]. The classification of orientably regular embeddings of complete graphs was initiated by N. L. Biggs [Big71] and completed by L. A. James and G. A. Jones [JaJ85]. Examples of latest progress in classification of orientably regular embeddings of complete bipartite graphs, complete multipartite graphs, and of cubes are due to G. A. Jones, R. Nedela and M. Škoviera [JNS04], J. H. Kwak and Y. S. Kwon [KwK05], and S. F. Du, J. H. Kwak and R. Nedela [DKN04, DKN05]. The phenomenon of chirality of maps is studied in great detail by A. Breda d'Azevedo, G. A. Jones, R. Nedela and M. Škoviera [BJN05]. The interested reader will find further information in references included in the papers listed.

3 Regular Maps, Groups, and Surfaces

We begin with giving details of the construction of regular maps from groups. Let G be a group generated by three involutions, two of which commute. In order to avoid entanglement into subtleties related to infinite degrees of vertices or faces of infinite length, assume that the product of any two generators of our group has a finite order. Then, G has a presentation of the form

$$G = \langle x, y, z |\ x^2 = y^2 = z^2 = (yz)^k = (zx)^m = (xy)^n = \ldots = 1 \rangle \qquad (1)$$

where dots indicate a possible presence of other relations. (As usual, in any such presentation we will assume that the exponents are *true* orders of the elements.) The regular map $M = M(G; x, y, z)$ that corresponds to (1) is constructed as follows. Consider, for each $g \in G$, a topological triangle labelled g, and label its sides with generators of G as in Fig. 1. The collection of all such triangles forms the set of *flags* of the map to be constructed. To simplify the matter, we will identify flags with their group labels. For each $g \in G$ and each $w \in \{x, y, z\}$, $w \neq 1$, we now identify the sides labelled w in the flags g and gw in such a way that the corresponding points where the thick, thin, or dashed sides meet are identified as well. This way we obtain a connected surface without boundary. The cellular decomposition of the surface induced by the union Γ of all thick segments forms our regular map $M = M(G; x, y, z)$. The 1-dimensional cell complex Γ is the *underlying graph* of the map. The identification of flags with elements of G was part of our construction. Other objects such as edges, vertices, and faces of the map M can be similarly identified with the left cosets of the subgroups $\langle x, y \rangle$, $\langle y, z \rangle$, and $\langle z, x \rangle$ in the group G, respectively, and their their mutual incidence is determined by non-empty intersection.

The two natural actions – by left and right multiplication – of the group G on the flag set of M (that is, on G itself) have important map-theoretical interpretations. For regular maps, right multiplication of flags $g \in G$ by the generators x, y, and z gives precisely the permutations X, Y and Z introduced in section 1. For the left multiplication, note that if two flags $h, h' \in G$ are related by some type of reflection, then for any $g \in G$ the flags gh and gh' are related by the same type of reflection. Left multiplication therefore preserves

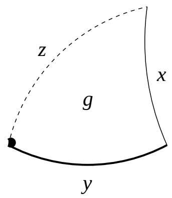


Fig. 1. A topological triangle representing a flag

the cell structure of M and induces an automorphism of M. This way the automorphism group $Aut(M)$ and its action on the map $M = M(G; x, y, z)$ may be identified with the group G and its action on itself by left multiplication. Consequently, regular maps of face length m and vertex valence k can be identified with presentations of finite 3-generator groups as in (1) where k and m represent *true orders* of the elements yz and zx. Briefly, in such a case we speak about regular maps of *type* $\{m, k\}$; listing the face length first is part of the traditional notation known as Schläfli symbol.

Features such as isomorphism and duality of regular maps can be conveniently explained with the help of our approach. Two regular maps $M_i = M(G_i; x_i, y_i, z_i)$, $i = 1, 2$, are *isomorphic* if there is a group isomorphism $\varphi : G_1 \to G_2$ such that $\varphi(x_1) = x_2$, $\varphi(y_1) = y_2$, and $\varphi(z_1) = z_2$. It is easy to check that this condition is equivalent to φ inducing an incidence preserving bijection between flags of M_1 and M_2. Setting $x' = y$ and $y' = x$ in the presentation (1), the map $M(G; x', y', z)$ formed from flags as in Fig. 1 but with thin side labelled $x' = y$ and thick side labelled $y' = x$ is the *dual map* of $M(G; x, y, z)$. The dual of a map M of type $\{m, k\}$ is usually denoted M^* and has type $\{k, m\}$. An illustration is in Fig. 2 which shows a regular embedding of the Petersen graph on the projective plane in solid lines and its dual regular map with underlying graph K_6 in dashed lines. Note that $(M^*)^* = M$ and that both M and M^* have the *same* (not merely isomorphic) automorphism groups. If M^* is isomorphic to M, then M is called *self-dual.*

In section 2 we noted that the theory of regular maps can be completely reduced to group theory. Indeed, it is obvious that instead of a regular map $M = M(G; x, y, z)$ we can just consider the group G *together* with its presentation (1); these will be referred to as $(k, m, 2)$*-groups.* Classification of regular maps of type $\{m, k\}$ up to isomorphism and duality is therefore equivalent to classification of $(k, m, 2)$-groups with $k \geq m$.

Consider now the two particular elements $r = yz$ and $s = zx$ of G. It can be checked that r and s represent a rotation of order k about the vertex and a rotation of order m about the centre of a face, both the vertex and the face centre being incident to the flag labelled 1. The subgroup $G^o = \langle r, s \rangle$ has index at most 2 in G, furnishing a simple orientablilty test: The supporting surface of the map is orientable if and only if the index $[G : G^\circ]$ is equal to 2. In general, if a map of type $\{m, k\}$ on an orientable surface contains two rotations r and s as described above, then the map is said to be *orientably*

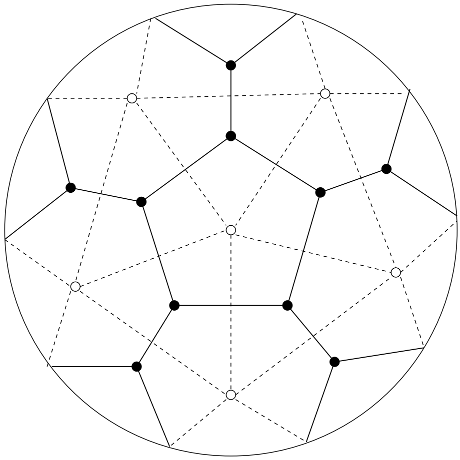


Fig. 2. A regular map and its dual in the projective plane

regular. Thus, every regular map on an orientable surface is orientably regular. Orientably regular maps that are not regular are called *chiral* in the literature. A prominent example of a chiral map is the (essentially, unique) triangular embedding of K_7 on the torus.

We assume familiarity with the concept of the Euler characteristic of a connected compact surface, which is equal to $2-2g$ or $2-h$ depending on whether the surface is orientable, of genus g, or nonorientable, of genus h. Any regular map M on a nonorientable surface of Euler characteristic χ has a natural double cover, the regular map $\tilde{M}$ on the corresponding orientable surface of Euler characteristic 2χ. Reversing the process, M arises from $\tilde{M}$ as a quotient by an antipodal reflection. At the algebraic level, an antipodal reflection of a map is simply a fixed-point-free orientation reversing automorphism commuting with all the orientation preserving automorphisms (and hence lying in the centre of the group, cf. [BeG89]). Such reflections are by no means unique in general. Different antipodal reflections applied to a regular map in the double cover may even yield non-isomorphic regular maps in the quotient surface, as was pointed out by S. Wilson [Wi78a].

The formalism introduced above enables us also to outline the links between the theory of regular maps, group theory, hyperbolic geometry, and complex functions. The extra notion we need is the one of the *full* $(k,m,2)$-*triangle group*, which is the group with presentation $\langle x,y,z |\ x^2=y^2=z^2=(yz)^k=(zx)^m=(xy)^2=1\rangle$. Because of absence of other relations, the corresponding regular map can be realised as a tessellation of a simply connected surface. This surface is the sphere, the euclidean plane, or the hyperbolic plane, depending on whether $1/k+1/m$ is greater than, equal to, or smaller than $1/2$. The tessellation is then formed by geometrically congruent regular m-gons, k of which meet at each vertex. The subgroup of all orientation-preserving automorphisms of this tessellation is the $(k, m, 2)$-*triangle group* $\langle r, s| \ r^k = s^m = (rs)^2 = 1\rangle$, which is an index-two subgroup of the full $(k, m, 2)$-triangle group obtained by letting $r = yz$ and $s = zx$.

The connections are now as follows. Except for embeddings of semi-stars in a sphere, automorphism groups of regular maps of a given type $\{m, k\}$ on compact surfaces are precisely the finite $(k, m, 2)$-groups, which are quotients of the full $(k, m, 2)$-triangle group by torsion-free normal subgroups of finite index. Likewise, finite orientably regular maps of type $\{m, k\}$ are in a one-to-one correspondence with normal, torsion-free, finite-index subgroups of the $(k, m, 2)$-triangle groups; if the subgroup is, at the same time, *not* normal in the *full* triangle group, the corresponding orientably regular map is *chiral*. These quotient constructions can be used to endow maps on compact surfaces with complex structure and geometry (spherical, euclidean, or hyperbolic). In particular, maps (not necessarily regular) can be regarded as complex algebraic curves over algebraic number fields. To conclude with a far-fetching connection, the algebraic curves view opens up a possibility to study the absolute Galois group by its action on maps, as suggested in the Grothendieck's programme [Gro84]. We recommend the survey papers by G. A. Jones [Jon97] and by G. A. Jones and D. Singerman [JoS96] for more details.

4 Regular Maps on Surfaces of Small Genus

In this section we survey regular maps on surfaces of orientable genus up to 15 and nonorientable genus up to 30. Until very recently, these were the only values of genera for which a classification of regular maps was known. In what follows we will be giving presentations of $(k, m, 2)$-groups in the form $\langle (x, y, z), \ \ldots\rangle$ where the (x, y, z) part will stand for $x, y, z|x^2 = y^2 = z^2 = (xy)^2$. For description of groups we will use the standard notation. That is, Z_n, D_n, S_n and A_n will denote the cyclic group of order n, the dihedral group of order $2n$, and the symmetric and the alternating groups of degree n, respectively.

Let G be a finite $(k, m, 2)$-group with a presentation of the form (1). The number of vertices, edges, and faces of the regular map $M = M(G; x, y, z)$ is simply obtained by dividing the number of flags, that is, the order $|G|$ of the group, by the orders of the dihedral stabilisers of the respective elements. The map M therefore has $v = |G|/(2k)$ vertices, $e = |G|/4$ edges, and $f = |G|/(2m)$ faces. By the Euler's formula we have $v - e + f = \chi$ where χ is the Euler characteristic of the supporting surface. Substituting for v, e, f gives

$$\chi = (1/k + 1/m - 1/2)|G|/2 \tag{2}$$

The surface with the largest Euler characteristic, 2, is the sphere. Regular maps on the sphere have been well known and their classification quickly follows from (2). It turns out that there are no chiral maps on the sphere.

Apart from embeddings of semi-stars (semi-edges incident to a single vertex), a spherical regular map arises either from an embedded k-cycle, or its dual (called a k-dipole), or one of the five Platonic polyhedra. We list the corresponding $(k, m, 2)$-groups $Aut(M)$ (which are all extended triangle groups since the sphere is simply connected) for the maps M up to duality, in the form $M/\mathrm{dual}(M)$:

embedded $k-$dipoles/$k-$cycles	$\langle(x, y, z),\ (yz)^k = (zx)^2 = 1\rangle \cong D_k \times Z_2$
tetrahedron (self $-$ dual)	$\langle(x, y, z),\ (yz)^3 = (zx)^3 = 1\rangle \cong S_4$
octahedron/cube	$\langle(x, y, z),\ (yz)^4 = (zx)^3 = 1\rangle \cong S_4 \times Z_2$
icosahedron/dodecahedron	$\langle(x, y, z),\ (yz)^5 = (zx)^3 = 1\rangle \cong A_5 \times Z_2$

The next simplest surface is the projective plane, which ia a nonorientable surface of Euler characteristic 1 (and of nonorientable genus 1). Since the sphere admits a unique antipodal reflection, regular maps on the projective plane are quotients of the spherical regular maps by the reflection. One has to be cautious, however, when looking for the antipodal reflection in terms of central involutions. For example, the groups G of spherical embeddings of k-dipoles always factor as $G^{\circ} \times \langle x\rangle$, but the automorphism induced by the left multiplication by x has fixed points; the same holds for k-cycles. A closer inspection shows that the spherical embeddings of odd cycles, odd dipoles, and the tetrahedron have no antipodal reflection. Using the notation $r = yz$ and $s = zx$, and taking duality into account, the unique antipodal reflection u in the remaining cases is given by $u = zsr^k$ for $2k$-dipoles, $u = zrs^{-1}r^2s$ for the octahedron, and $u = zr^2sr^{-1}sr^{-2}s$ for the icosahedron. Since the antipodal reflection is central, we have $G = \langle x, y, z\rangle \cong \langle r, s\rangle \times \langle u\rangle \cong G^{\circ} \times Z_2$ and the quotients arise by dividing out by $\langle u\rangle$. Antipodal quotients of our spherical maps are therefore the projective-planar embeddings of k-cycles and their duals (bouquets of k circles), K_4 and its dual (denoted by $K_3^{(2)}$, which is K_3 with doubled edges), and K_6 and its dual (formed by the Petersen graph P embedded as in Fig. 2). Presentations of the corresponding groups $Aut(M)$, again in the form $M/\mathrm{dual}(M)$, are as follows (using $r = yz$ and $s = zx$):

embedded bouquets/cycles	$\langle(x, y, z),\ r^{2k} = s^2 = zsr^k = 1\rangle \cong D_{2k}$
embedded $K_3^{(2)}$/embedded K_4	$\langle(x, y, z),\ r^4 = s^3 = zrs^{-1}r^2s = 1\rangle \cong S_4$
embedded K_6/embedded P	$\langle(x, y, z),\ r^5 = s^3 = zr^2sr^{-1}sr^{-2}s = 1\rangle \cong A_5$

Note that in contrast with the spherical case, presentations of the groups of the projective-planar regular maps contain an extra relator coming from the antipodal involution u. Extra relators feature in all presentations of regular maps on non-simply connected surfaces, since they reflect presence of non-contractible curves.

Classification of regular and chiral maps on the torus (an orientable surface of genus 1, with Euler characteristic 0) was initiated by H. Brahana [Bra26] and all details were eventually supplied by H. S. M. Coxeter [Cox48]. Euler's formula tells us that a toroidal regular map of type $\{m, k\}$, $m \le k$, can exist only if either $k = m = 4$, or $k = 6$ and $m = 3$. By the theory outlined in section 3, for the classification it is sufficient to identify all finite-index, torsion-free, normal subgroups of the $(4, 4, 2)$- and $(6, 3, 2)$-triangle groups and of the related full triangle groups. These are groups of (possibly orientation-reversing) euclidean isometries, leaving invariant the corresponding tessellation. Using representations of the four groups either by matrices or by complex numbers it can be shown that all their finite-index torsion-free normal subgroups are generated by powers of two specific commuting translations. As the result, toroidal chiral and regular maps can be parametrised by ordered pairs of integers (b, c) representing powers of the two translations. Up to isomorphism and duality, the automorphism groups of all *chiral* toroidal maps satisfy $bc(b - c) \neq 0$ and have presentations of the form

$$\text{Type } \{4,4\} : \langle r, s \,|\, r^4 = s^4 = (rs)^2 = (rs^{-1})^b (r^{-1}s)^c = 1 \rangle$$
$$\text{Type } \{6,3\} : \langle r, s \,|\, r^6 = s^3 = (rs)^2 = (rs^{-1}r)^b (s^{-1}r^2)^c = 1 \rangle$$

In the case when $bc(b - c) = 0$ we obtain regular toroidal maps. Using $r = yz$ and $s = zx$, their groups have presentations

$$\text{Type } \{4,4\} : \langle (x, y, z), r^4 = s^4 = (rs^{-1})^b (r^{-1}s)^c = 1 \rangle$$
$$\text{Type } \{6,3\} : \langle (x, y, z), r^6 = s^3 = (rs^{-1}r)^b (s^{-1}r^2)^c = 1 \rangle$$

Unlike spherical maps, toroidal regular maps do not admit antipodal reflections. Consequently, there are no regular maps on the Klein bottle, the nonorientable surface of Euler characteristic 0 that is double-covered by the torus. Another way to see this is to invoke the known fact that the only automorphism of the Klein bottle acting as a rotation about some point must have order two. This completes the classification of regular maps on surfaces of non-negative Euler characteristic. Note that while the number of regular maps on the sphere, on the projective plane, and on the torus is infinite, the infinitude in the first two cases is due to rather trivial maps.

In contrast with this, the number of regular maps on any compact surface $\mathcal{S}$ of negative Euler characteristic is *finite*. Indeed, from (2) it is obvious that if $\chi(\mathcal{S}) < 0$, then $1/k + 1/m < 1/2$ and $|G| = -2\chi(\mathcal{S})/(1/2 - 1/k - 1/m)$. Among all pairs (k, m) for which $1/k + 1/m < 1/2$, the reciprocal of the denominator achieves the largest value, 42, precisely when $\{k, m\} = \{3, 7\}$. We thus arrive at the *Hurwitz bound* $|G| \leq -84\chi(\mathcal{S})$ if $\chi(\mathcal{S}) < 0$, giving a cap on the order of the automorphism group of a regular map on a surface with negative Euler characteristic. Similarly, the order of the automorphism group of a chiral map on an orientable surface of genus $g \geq 2$ cannot exceed $84(g - 1)$. The bound is named after H. Hurwitz who first proved its orientable variant for groups of conformal automorphisms acting on Riemann surfaces; see T. W. Tucker [Tuc83].

To classify regular maps on a given surface of negative Euler characteristic, one could in principle use the strategy outlined in section 3: Work out the admissible types $\{m, k\}$ from the Euler's formula (2) and then determine the torsion-free normal subgroups of the corresponding $(k, m, 2)$-triangle groups of index not exceeding the Hurwitz bound. The second step in full generality, however, seems to be far beyond the reach of the currently available methods. If $1/k + 1/m < 1/2$, that is, when the type $\{m, k\}$ is *hyperbolic*, the full $(k, m, 2)$-triangle group is a subgroup of the group of (direct as well as orientation reversing) hyperbolic isometries, leaving invariant a regular tessellation of the hyperbolic plane of type $\{m, k\}$. As opposed to the euclidean case, very little is known about normal subgroups of such hyperbolic triangle groups.

How can one then approach the problem? In the early stages, a number of results were obtained by relation-chasing, that is, trying to determine the extra relations one has to add in the presentation of a triangle group or a full triangle group to obtain a quotient group and a quotient map on a given surface. Combined with other known facts and methods, mostly of group-theoretical nature, a classification for orientable surfaces of Euler characteristic $\chi = -2$ and -4 (and hence genus 2 and 3) was given by H. S. M. Coxeter and W. O. J. Moser [CMo84] and F. A. Sherk [She59], respectively. None of the maps for the two genera are chiral. Also, regular maps of genus 2 turn out to have no antipodal reflections, implying that there are no regular maps on the nonorientable surface with $\chi = -1$. Using similar methods, A. S. Grek in a series of papers [Gre63, Gr66a, Gr66b] derived a classification of regular maps on non-orientable surfaces with $-2 \geq \chi \geq -4$.

A more powerful method based on permutation representations is due to D. Garbe [Gar69], introduced in the course of classification of regular maps on the orientable surface of genus 4. Suppose that one wants to classify regular maps of type $\{m, k\}$ with exactly d faces. This is equivalent to classifying all $(k, m, 2)$-groups G of order $2md$. Since G contains a dihedral subgroup H of order $2m$, we have a permutation representation of G of degree d given by the action of G on the cosets of H; the image of H in this representation is the stabiliser of an element. Now, let $T = \langle (x, y, z), (yz)^k = (zx)^m = 1 \rangle$ be the full $(k, m, 2)$-triangle group and let N be the torsion-free normal subgroup of T such that $T/N \cong G$. Because of absence of torsion, the image of the dihedral group $L = \langle z, x \rangle$ under the natural projection $\theta : T \to T/N$ is again a dihedral group of order $2m$; in fact, we may assume that the image is H. We do not know G and N yet, but observe that the transitive permutation representation of G of degree d lifts onto a transitive permutation representation of T of the same degree d but this time on the cosets of the subgroup $\theta^{-1}(H)$. This suggests the following algorithm to determine all such groups G.

(A) Construct all transitive permutation representations $\psi : T \to S_d$ where S_d is the symmetric group of degree d acting on the set $\{1, 2, \ldots, d\}$, such that $\psi(L) = \mathrm{Stab}(1)$, the stabiliser of the element 1 in the image $\psi(T)$.

(B) Construct all epimorphisms ϑ from the subgroup $K = \psi^{-1}(\mathrm{Stab}(1)) < T$ onto the dihedral group D_m of order $2m$, such that $\vartheta(L) = D_m$ and such that $N = \ker(\vartheta)$ is torsion-free.

Then, clearly, $G = T/N$ is a $(k, m, 2)$-group and the corresponding regular map has exactly d faces, that is, $|G| = 2md$. Indeed, an easy calculation shows that $|G| = |T/N| = [T : N] = [T : K][K : N] = [\psi(T) : \psi(K)]|D_m| = [\psi(T) : \mathrm{Stab}(1)] \cdot 2m = 2md$, as claimed. The fact that this algorithm constructs *all* regular maps of type $\{m, k\}$ with exactly d faces follows from the above discussion. In addition, all such *orientable* maps are filtered out by the condition that N be a subgroup of the orientation preserving part of T, that is, $N < \langle r, s \rangle$ where $r = yz$ and $s = zx$. Running the procedure with T, L, and D_m replaced by the $(k, m, 2)$-triangle group $\langle r, s |\ r^k = s^m = (rs)^2 = 1 \rangle$, the cyclic group $\langle s \rangle$, and the cyclic group Z_m, respectively, and identifying the normal subgroups N of T that are *not* normal in the *full* $(k, m, 2)$-triangle group, one obtains all the chiral maps with d faces.

In practice, for each permutation representation in (A) one finds with the help of the Reidemeister-Schreier method a presentation of K and then one searches over the epimorphisms in (B). As long as the number d of faces is relatively small, the calculations – although time consuming and far from trivial – can be done by hand. This was the main tool used in the classification of regular and chiral maps on orientable surfaces of genus 5 and 6 (P. Bergau and D. Garbe [BeG89]), and 7 (D. Garbe in [Gar78]). The corresponding nonorientable results were obtained for Euler characteristic -5 by J. Scherwa [Sch85] and for -6 by P. Bergau and D. Garbe [BeG89]. In this connection it is worth noting that S. E. Wilson [Wi78b] has outlined a similar algorithm using a geometric language.

Summing up, by the late 1980's, the collective effort of the researchers mentioned above resulted in classification of all regular and chiral maps on orientable surfaces of Euler characteristic $\chi \geq -12$ (that is, up to genus 7), and regular maps on nonorientable surfaces with $\chi \geq -6$ (up to genus 8). It is interesting to note that, in the orientable case, there are no chiral maps of orientable genus between 2 and 6 at all! Further progress came in about a decade, when M. Conder and P. Dobcsányi [CoD01] published a computer-assisted classification of all regular and chiral maps on orientable surfaces up to genus 15, and regular maps on nonorientable surfaces up to genus 30. The authors used their own adaptation of the low-index subgroup algorithm and applied it to finding 'small' index normal subgroups – but not subgroups of the full triangle groups. Instead, it turned out to be of advantage to consider normal subgroups of the group $\langle x, y, z |\ x^2 = y^2 = z^2 = (xy)^2 = 1 \rangle$, with the relator $(zx)^3$ or $(zx)^4$ added in the case when $m = 3$ or $m = 4$, and extract the rest from there. Re-confirming all the earlier results, the list in [CoD01] documents the state-of -the-art of the regular maps classification problem at the end of the millenium. In particular, by then a complete classification was known only for a finite number of surfaces.

5 Regular Maps on Surfaces of Large Genus

A breakthrough in classification of regular maps was achieved when A. Breda d'Azevedo, R. Nedela and the author [BNS05] classified regular maps on *all* surfaces of negative prime Euler characteristic – that is, on an *infinite* number of surfaces. Let $n(p)$ denote the number of regular maps, up to isomorphism and duality, on a surface with Euler characteristic $-p$, where p is a prime. The numbers $n(p)$ for $p < 29$ together with the automorphism groups of the maps have been determined by the computer-aided classification of [CoD01] and we therefore state the result of [BNS05] for the larger primes only. Also, for $p \equiv -1 \pmod 4$ let $\nu(p)$ be the number of pairs (j,l) such that $j > l \geq 3$, both j and l are odd, coprime, and $(j-1)(l-1) = p+1$.

Theorem 5.1 ([BNS05]). *Let p be an odd prime, $p \geq 29$ Then, up to isomorphism and duality, for the number $n(p)$ of regular maps on a surface with Euler characteristic $-p$ and for the corresponding groups G we have:*

(A) $n(p) = 0$ *if* $p \equiv 1 \pmod{12}$*;*
(B) $n(p) = 1$ *if* $p \equiv 5 \pmod{12}$*, with* $G \cong G_p = \langle r, s |\ r^{p+4} = s^4 = (rs)^2 = sr^3s^{-1}r^3 = 1\rangle$*;*
(C) $n(p) = \nu(p)$ *if* $p \equiv 7 \pmod{12}$*, with* $G \cong G_{j,l} = \langle r, s |\ r^{2j} = s^{2l} = (rs)^2 = (rs^{-1})^2 = 1\rangle$*;*
(D) $n(p) = \nu(p) + 1$ *if* $p \equiv -1 \pmod{12}$*; the groups G are G_p and the $\nu(p)$ groups $G_{j,l}$.*

Presentations of the above groups are given in terms of $r = yz$ and $s = zx$ only; recall that for nonorientable maps the automorphism group os generated by r and s. This implicitly assumes that the involutions x, y, z can be recovered from r and s in a unique way. While this is not true in general, in our case it can be shown that the involutions are indeed unique. We note that the group G_p is an extension of $Z_{(p+4)/3}$ by S_4 while for the groups $G_{j,l}$ we have $G_{j,l} \simeq D_j \times D_l$.

An important ingredient of the proof of Theorem 5.1 is the following milestone result in the project of classification of finite simple groups, due to D. Gorenstein and J.H. Walter [GoW65]: *If H is a group with a dihedral Sylow 2-subgroup and if $O(H)$ is the (unique) maximal normal subgroup of G of odd order, then $H/O(H)$ is isomorphic to either a Sylow 2-subgroup of G, or to the alternating group A_7, or to a subgroup of* $\mathrm{Aut}(PSL(2,q))$ *containing $PSL(2,q)$, where q is and odd prime power.* The original proof has about 180 pages and depends on the celebrated Feit-Thompson odd-order groups theorem. Subsequently, a shorter (but by no means simpler) proof of the Gorenstein-Walter result was given in [Ben81, BGl81] but it still depends on the odd-order groups theorem. An interesting challenge would be to find a proof of Theorem 5.1 without invoking such high-calibre results.

Every orientable surface carries a regular map; an example for any $g \geq 1$ is a single-face regular embedding of a bouquet of $2g$ circles. While M. Conder and B. Everitt [CoE95] showed that more than three quarters of nonorientable surfaces support a regular map, there are "gaps". Absence of regular maps on nonorientable surfaces with Euler characteristic $\chi = 0$ (the Klein bottle) and $\chi = -1$ has been known for a long time. The list of [CoD01] shows that there are no regular maps on nonorientable surfaces with $\chi = -16$, -22, and -25, either.

From a deep study by S. Wilson and A. Breda [WBr04] it follows that nonexistence of regular maps extends also to $\chi = -37$ and -46 and that all the above are the only gaps in the range $-50 \leq \chi \leq 0$. The amazing consequence of Theorem 5.1 is that *there are infinitely many gaps*! Specifically, part (A) of the theorem and the list of [CoD01] imply that there are no regular maps on surfaces of Euler characteristic $\chi = -p$ where p is a prime congruent to 1 (mod 12) and $p \neq 13$.

Although a number of steps in the proof of Theorem 5.1 substantially depended on the primality of $-\chi$, extensions to small odd multiples of primes are within reach. A very recent work by G. A. Jones, R. Nedela and the author [JNS05] has resulted in a classification of regular maps on (nonorientable) surfaces with $\chi = -3p$ for sufficiently large primes.

A complete classification of regular or chiral maps on an infinite family of *orientable* surfaces is still not available. Nevertheless, important contributions to the study of automorphism groups of Riemann surfaces by M. Belolipetsky and G. A. Jones [BeJ04] imply ingredients for a classification of the regular and chiral maps on orientable surfaces of genus $p+1$ with 'large' automrophism group G in the sense that $|G| > 12p$ for the regular case, and $|G| > 6p$ for the chiral case, where p is a prime. In order to see what this means for regular maps, observe that the Euler characteristic of such a surface is $\chi = -2p$. From (2) we then obtain $|G| = -2\chi/(1/2 - 1/k - 1/m) > -4\chi = 8p$. In general, however, one may have regular maps with groups of order between $8p$ and $12p$ that are not captured by the results of [BeJ04].

A fair amount of structural information about automorphism groups of regular maps on a given surface can be extracted from Euler's formula (2) by just arithmetic considerations. An example is the following lemma due to M. Conder, P. Potočnik and the author [CPS04], which overlaps with an observation made by G. A. Jones [Jon04].

Lemma 5.2. *Let G be the automorphism group of a regular map on a surface with Euler characteristic $\chi \neq 0$ and let p' be a prime divisor of $|G|$ coprime with χ. Then, the Sylow p'-subgroups of G are cyclic if p' is odd, or dihedral if $p' = 2$. In particular, the Sylow 2-subgroups of G are automatically dihedral if χ is odd.*

This opens up the possibility of application of the Gorenstein-Walter theorem in the study of regular maps on surfaces with any odd Euler characteristic (which are necessarily nonorientable). Unfortunately, one still needs more information to determine the exact structure of the groups. Inspired by the first part of Lemma 5.2 it makes sense to look for cyclicity of the odd-order Sylow subgroups. And indeed, we have a good example at hand: Revisiting the proof of Theorem 5.1, it is easy to see that the claim (i) immediately implies that *all* the odd-order Sylow subgroups of the groups of regular maps on surfaces of Euler characteristic $\chi = -p$, p an odd prime, are cyclic! The same happens to be true for the groups of regular maps on surfaces with $\chi = -3p$ except for the groups $D_j \times D_l$ when $(j,l) = 3$. Another substantial contribution here is a further result of [CPS04]:

Proposition 5.3. *Let G be the automorphism group of a regular map on a surface with Euler characteristic $-p^2$ where p is an odd prime and $p \neq 3$. Then, every odd-order Sylow subgroup of G is cyclic.*

We see that there is more than enough motivation to investigate $(k,m,2)$-groups G such that the Sylow 2-subgroups of G are dihedral and *all* the odd-order Sylow subgroups of G are cyclic. Such a class of groups in general has appeared in the literature in connection with various questions in group theory. Determination of groups with the property that all its Sylow subgroups (including the Sylow 2-subgroups) are cyclic goes back to W. Burnside and even to G. Frobenius (see [Bur11]). Relaxations of the Sylow 2-subgroups condition, however, turns the problem to a challenge which has been resolved only in special cases. We will say that H is a group of *Zassenhaus type* if all the odd-order Sylow subgroups of H are cyclic and if every Sylow 2-subgroup of H contains a cyclic subgroup of index 2. All the solvable groups of Zassenhaus type were determined by H. Zassenhaus [Zas36] and the characterisation of all such non-solvable groups was completed by M. Suzuki [Suz55] and W. J. Wong [Won66].

If G is a $(k,m,2)$-group with presentation (1), the regular map $M = M(G;x,y,z)$ will be called a map of *Zassenhaus type* if the group G is of Zassenhaus type. In a recent work of M. Conder, P. Potočnik and the author, the above results together with the Gorenstein-Walter theorem have been used to classify all regular maps of Zassenhaus type. Since details of the classification are a little too long to be reproduced here, we just give a brief summary about the groups. In the case when the group G is solvable, the maps fall into 8 classes and G is either a dihedral group, or a split extension of a cyclic group by the Klein four-group, or else a split extension of the Klein four-group by a dihedral group (with certain congruence restrictions on the orders). If M is a map of Zassenhaus type with a non-solvable group G, then either $G \cong PGL(2,q)$ for some prime q (not a non-trivial power of a prime), or G is an extension of such a $PGL(2,q)$ by a cyclic group of order coprime with $q(q^2-1)$. As an aside, regular maps with automorphism groups isomorphic to projective (special as well as general) linear two-dimensional groups have been classified in great detail by the same set of authors in [CPS05].

These results are solid tools for investigation of regular maps of a given type $\{m,k\}$ on a surface with a given Euler characteristic χ. A nice corollary is a classification of regular maps of type $\{m,k\}$ on surfaces with $\chi = -p^2$

where p is a prime not dividing both k and m, such that $p \geq 23$; we refer to [CPS04] for particulars.

6 Conclusion: Regular Hypermaps

A natural avenue of research is to extend the classification to hypermaps. At the group-theoretic level, a regular (k, m, l)-*hypermap* is simply a finite group G generated by three involutions x, y, z in which the products yz, zx, and xy have orders k, m and l, respectively. Regular maps are therefore a special case of regular (k, m, l)-hypermaps for $l = 2$. Orientably regular and chiral hypermaps are defined analogously as in the case of maps. The theory of regular and chiral hypermaps is similar to the theory of regular and chiral maps, including connections with hyperbolic geometry and complex functions. Pictorial representations, however, are a little different and not unique. A topological representation of a hypermap that gives equal status to hypervertices, hyperedges, and hyperfaces (reflecting the equal status of the products yz, zx, and xy in the corresponding triangle group) can be obtained by means of an embedding of a trivalent graphs with faces of lengths $2k$, $2m$ and $2l$, alternating at each vertex.

The classification of regular maps on surfaces of negative prime Euler characteristic [BNS05] was extended to regular hypermaps by G. A. Jones [Jon04]. It turns out that the classification is obtained from the list of [BNS05] by adding just two items, the regular $(4, 3, 3)$- and $(6, 5, 4)$-hypermaps with groups $PSL(2, 7)$ and $PGL(2, 5) \cong S_5$. It is also worth mentioning that the deep study of gaps in the nonorientable range of Euler characteristics χ for $-50 \leq \chi < 0$ by S. Wilson and A. Breda [WBr04], which we have mentioned in the previous section, was actually done for regular hypermaps in general.

A combination of the approaches outlined in this brief survey, perhaps blended with the theory of linear and permutation representations of groups and with the general knowledge on group actions on compact surfaces, is likely to yield substantial contributions to the theory regular maps on a fixed surface in the indicated directions.

References

[BeJ04] M. Belolipetsky, G. A. Jones, Automorphism groups of Riemann surfaces of genus $p + 1$, where p is a prime, Submitted (2004).

[Ben81] H. Bender, Finite groups with dihedral Sylow 2-subgroups, *J. of Algebra* **70** (1981), 216–228.

[BGl81] H. Bender, G. Glauberman, Characters of finite groups with dihedral Sylow 2-subgroups, *J. Algebra* **70** (1981), 200–215.

[BeG89] P. Bergau, D. Garbe, Non-orientable and orientable regular maps, in *Proceedings of "Groups-Korea 1988"*, *Lect. Notes Math.* **1398**, Springer (1989), 29–42.

[Big71] N. L. Biggs, Automorphisms of imbedded graphs, *J. Combinat. Theory Ser. B* **11** (1971), 132–138.
[Bra26] H. R. Brahana, Regular maps on an anchor ring, *Amer. J. Math.* **48** (1926), 225–240.
[Bra27] H. R. Brahana, Regular maps and their groups, *Amer. J. Math.* **49** (1927), 268–284.
[BJN05] A. Breda d'Azevedo, G. A. Jones, R. Nedela, M. Škoviera, Chirality groups of maps and hypermaps, submitted (2005).
[BNS05] A. Breda d'Azevedo, R. Nedela, J. Širáň, Classification of regular maps of negative prime Euler characteristic, *Trans. Amer. Math. Soc.* **357** No. 10 (2005), 4175–4190.
[BrS85] R. P. Bryant, D. Singerman, Foundations of the theory of maps on surfaces with boundary, *Quart. J. Math. Oxford* Ser. (2) **36** No. 141 (1985), 17–41.
[Bur11] W. Burnside, *Theory of Groups of Finite Order*, Cambridge Univ. Press, 1911.
[CoD01] M. D. E. Conder, P. Dobcsányi, Determination of all regular maps of small genus, *J. Combinat. Theory Ser. B* **81** (2001), 224–242.
[CoE95] M. D. E. Conder, B. Everitt, Regular maps on non-orientable surfaces, *Geom. Dedicata* **56** (1995), 209–219.
[CPS04] M. D. E. Conder, P. Potočnik, J. Širáň, Regular maps on groups with cyclic odd-Sylow subgroups, preprint (2004).
[CPS05] M. D. E. Conder, P. Potočnik, J. Širáň, Regular hypermaps over projective linear groups, preprint (2005).
[Cox48] H. S. M. Coxeter, Configurations and maps, *Rep. Math. Colloq (2)* **8** (1948), 18–38.
[CMo84] H. S. M. Coxeter, W. O. J. Moser, *Generators and Relations for Discrete Groups*, 4th Ed., Springer-Verlag, Berlin 1984.
[DKN04] S. F. Du, J. H. Kwak, R. Nedela, Regular embeddings of complete multipartite graphs, submitted (2004).
[DKN05] S. F. Du, J. H. Kwak, R. Nedela, Classification of regular embeddings of hypercubes of odd dimension, submitted (2005).
[Dyc80] W. Dyck, Über Aufstellung und Untersuchung von Gruppe und Irrationalität regularer Riemannscher Flächen, *Math. Ann.* **17** (1880), 473–508.
[Gar69] D. Garbe, Über die regulären Zerlegungen geschlossener orientierbarer Flächen, *J. Reine Angew. Math.* **237** (1969), 39–55.
[Gar78] D. Garbe, A remark on non-symmetric Riemann surfaces, *Arch. Math.* **30** (1978), 435–437.
[GNS99] A. Gardiner, R. Nedela, J. Širáň, M. Škoviera, Characterization of graphs which underlie regular maps on closed surfaces, *J. London Math. Soc. (2)* **59** No. 1 (1999), 100–108.
[GoW65] D. Gorenstein, J. H. Walter, The characterization of finite groups with dihedral Sylow 2-subgroups, I, II, III, *J. of Algebra* **2** (1965), 85–151, 218–270, 334–393.
[Gre63] A. S. Grek, Reqular polyhedra of simplest hyperbolic types, (Russian) *Ivanov. Gos. Ped. Inst. Učen. Zap.* **34** (1963), 27–30.
[Gr66a] A. S. Grek, Regular polyhedrons on surfaces with Euler characteristic $\chi = -4$, (Russian) *Soobšč. Akad. Nauk Gruzin. SSR* **42** (1966), 11–15.
[Gr66b] A. S. Grek, Regular polyhedra on a closed surface with the Euler characteristic $\chi = -3$, (Russian) *Izv. Vysš. Učebn. Zaved. Matematika* **55** No. 6 (1966), 50–53.

[Gro84] A. Grothendieck, Esquisse d'un programme, preprint, Montpellier (1984).
[GrT87] J.L. Gross, T.W. Tucker, *Topological Graph Theory*, Wiley (1987) and Dover (2001).
[Hea90] P.J. Heawood, Map-colour theorem. Quart, *J. Math.* **24** (1890), 332–338.
[Hef98] L. Heffter, Über metazyklische Gruppen und Nachbarconfigurationen, *Math. Ann.* **50** (1898), 261–268.
[JaJ85] L.D. James, G.A. Jones, Regular orientable imbeddings of complete graphs, *J. Combinat. Theory Ser. B* **39** (1985), 353–367.
[Jon97] G.A. Jones, Maps on surfaces and Galois groups, *Math. Slovaca* **47** (1997), 1–33.
[Jon04] G.A. Jones, Large groups of automorphisms of compact surfaces with prime negative Euler characteristic, preprint (2004).
[JNS05] G.A. Jones, R. Nedela, J. Širáň, Classification of regular maps of Euler characteristic $-3p$, where p is a prime, preprint (2005).
[JNS04] G.A. Jones, R. Nedela, M. Škoviera, Regular embeddings of $K_{n,n}$ where n is odd prime power, preprint (2004).
[JoS78] G.A. Jones, D. Singerman, Theory of maps on orientable surfaces, *Proc. London Math. Soc.* (3) **37** (1978), 273–307.
[JoS96] G.A. Jones, D. Singerman, Belyĭ functions, hypermaps, and Galois groups, *Bull. London Math. Soc.* **28** (1996), 561–590.
[Kep19] J. Kepler, *The harmony of the world*, (Translation from the Latin "Harmonice Mundi", 1619) *Memoirs Amer. Philos. Soc.* **209**, American Philosophical Society, Philadelphia, PA (1997).
[Kle79] F. Klein, Über die Transformation siebenter Ordnung der elliptischen Functionen, *Math. Ann.* **14** (1879), 428–471.
[KwK05] J.H. Kwak, Y.S. Kwon, Regular orientable embeddings of complete bipartite graphs, preprint (2005).
[Sah69] C.H. Sah, Groups related to compact Riemann surfaces, *Acta Math.* **123** (1969), 13–42.
[Sch85] J. Scherwa, *Regulaere Karten geschlossener nichtorientierbarer Flaechen*, Diploma Thesis, Bielefeld (1985).
[She59] F.A. Sherk, The regular maps on a surface of genus three, *Canad. J. Math.* **11** (1959), 452–480.
[Suz55] M. Suzuki, On finite groups with cyclic Sylow subgroups for all odd primes, *Amer. J. Math.* **77** (1955), 657–691.
[Tuc83] T.W. Tucker, Finite groups acting on surfaces and the genus of a group, *J. Combinat. Theory Ser. B* **34** No. 1 (1983), 82–98.
[Tut73] W.T. Tutte, What is a map? In *New Directions in Graph Theory* (F. Harary, Ed.), Acad. Press (1973), 309–325.
[Wi78a] S.E. Wilson, Non-orientable regular maps, *Ars Combin.* **5** (1978), 213–218.
[Wi78b] S.E. Wilson, Riemann surfaces over regular maps, *Canad. J. Math.* **30** (1978), 763–782.
[WBr04] S.E. Wilson, A. Breda D'Azevedo, Surfaces having no regular hypermaps, *Discrete Math.* **277** (2004), 241–274.
[Won66] W.J. Wong, On finite groups with semi-dihedral Sylow 2-subgroups, *J. Algebra* **4** (1966), 52–63.
[Zas36] H. Zassenhaus, Über endliche Fastkörper, *Abh. Math. Sem. Univ. Hamburg* **11** (1936), 187–220.

Part VII

On Six Problems Posed by Jarik Nešetřil

Jørgen Bang-Jensen[1], Bruce Reed[2], Mathias Schacht[3⋆], Robert Šámal[4], Bjarne Toft[5], and Uli Wagner[6⋆⋆]

[1] Dept. of Mathematics & Computer Science, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark
jbj@imada.sdu.dk
[2] School of Computer Science, McGill University, 3480 University, Montreal, Quebec, H3A 2A7 Canada
breed@cs.mcgill.ca
[3] Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany
schacht@informatik.hu-berlin.de
[4] Charles University, Faculty of Mathematics and Physics, Institute for Theoretical Computer Science (ITI), Malostranské náměstí 25, 118 00 Prague, Czech Republic
samal@kam.mff.cuni.cz
[5] Department of Mathematics and Computer Science, The University of Southern Denmark, DK-5230 Odense, Denmark
btoft@imada.sdu.dk
[6] Einstein Institute of Mathematics, Edmond J. Safra Campus, Givat Ram, The Hebrew University of Jerusalem, Jerusalem, 91904 Israel
uli@math.huji.ac.il

In this article we present several open problems posed (or co-posed) by Jarik Nešetřil. The choice was guided by two criteria. First, we restricted ourselves to problems that are simple to state and (therefore) possible to explain to a non-specialist in the given field. Second, we selected problems that, while being still open, did stimulate research by other people and have an interesting development behind them.

Most of the problems seem to be of fundamental nature and central. For all of them a simple argument might possibly solve them, yet this argument has eluded many researchers for many years. The inspiration from Jarik Nešetřil over the years is much appreciated!

⋆ The author was supported by DFG grant SCHA 1263/1-1.
⋆⋆ Partially supported by the Edmund Landau Center for Research in Mathematical Analysis, supported by the Minerva Foundation (Germany).

1 Minimally Asymmetric Oriented Graphs

Let G be an oriented graph. Suppose G is asymmetric, but every vertex-deleted subgraph $G - v$ fails to be asymmetric. Is it true that G must be K_1?

Jaroslav Nešetřil, Oberwolfach seminar, 1988

A (di)graph is *asymmetric* if its automorphism group is trivial, that is, contains only the identity element. A (di)graph is *symmetric* if it has at least one non-trivial automorphism. An asymmetric (di)graph G is *minimally asymmetric* if G is asymmetric, but $G - v$ is symmetric for every vertex v of G.

The problem was posed by Nešetřil at several conferences and according to Wójcik [Wój96] at least as early as during the Oberwolfach seminar in 1988. It is probably even older than that. At the Oberwolfach seminar in 1988 Nešetřil also conjectured that there are only finitely many minimally asymmetric undirected graphs.

In [NS92] and [Sab91] undirected minimally asymmetric graphs are studied. It turns out [Sab91] that a useful property to use in this context is the length λ of a longest induced path. It is shown in that paper that there are no minimally asymmetric graphs with $\lambda \geq 6$ and precisely two minimally asymmetric graphs with $\lambda = 5$. In [NS92] it is shown that there are exactly seven finite minimal asymmetric graphs with $\lambda = 4$.

Clearly, for every minimally asymmetric graph G one obtains a minimally asymmetric digraph D by replacing each edge of G by a directed 2-cycle. An oriented graph is a digraph obtained from a graph by orienting each of its edges. In particular an oriented cycle is obtained form an undirected cycle in this way. In [Wój96] Wójcik proved the following result. Recall that a cycle is symmetric if it has a non-trivial automorphism.

Theorem 1.1. *Every minimally asymmetric digraph contains a symmetric cycle.*

This implies in particular that there are no minimally asymmetric trees (a fact also proved earlier by Nešetřil) and that the conjecture holds for many classes of asymmetric acyclic digraphs. One example is a transitive tournament. Note that acyclic digraphs may contain symmetric cycles (e.g. $1 \to 2 \leftarrow 3 \to 4 \leftarrow 1$ which has a non-trivial automorphism without fixpoints) so Theorem 1.1 does not immediately seem to imply that there are no minimally asymmetric acyclic digraphs.

References

[Neš71] J. Nešetřil, A congruence theorem for asymmetric trees, *Pacific Journal of Mathematics* **37**, 771–778.

[Neš99] J. Nešetřil, Asymmetric graphs and colouring problems, Problems presented at the 6th midsummer combinatorial workshop, Prague 1999.
[NS92] J. Nešetřil and G. Sabidussi, Minimal asymmetric graphs of induced length 4, *Graphs and Combinatorics* **8** (1992), 343–359.
[Sab91] G. Sabidussi, Clumps, minimal asymmetric graphs, and involutions, *J. Combin. Th. Ser. B* **53** (1991), 40–79.
[Wój96] P. Wójcik, On automorphisms of digraphs without symmetric cycles, *Comment. Math. Univ. Carolinae* **37** (1996), 457–467.

2 Partition into Induced Matchings alias The Strong Chromatic Index

The edges of a graph G of maximum degree Δ can be partitioned into at most $\frac{5}{4}\Delta^2$ colour classes, each of which induces a matching.

Paul Erdős and Jaroslav Nešetřil, a combinatorial seminar at Charles University, Prague, 1985

This conjecture was made by Erdős and Nešetřil at a combinatorial seminar at Charles University in Prague in 1985. One year later it was presented at Colloquium on Irregularities of Partitions in Fertőd, Hungary (see [EN89]). For every graph H, the chromatic number of H is at most $\Delta(H)+1$ and the chromatic number of its square is at most $\Delta(H)^2+1$. Vizing's theorem tells us that for line graphs, we can improve the first result, essentially by a factor of 2. The conjecture above suggests that a similar improvement is possible for the second result.

Erdős and Nešetřil had noticed that if we take a cycle of length five and replace each vertex by a stable set of size k, joining two new vertices precisely if the corresponding two vertices of the five cycle are adjacent, then the square of the line graph of the resultant graph is a clique. This shows that the above conjecture is tight when Δ is even. For odd Δ, Erdős and Nešetřil actually made the stronger conjecture that the chromatic number of the square of a line graph of maximum degree Δ is at most $\frac{5}{4}\Delta^2-\frac{\Delta}{2}+\frac{1}{4}$, which again is tight because of a similar example.

The conjecture was proven for $\Delta=3$ by Andersen [And92] and independently Horak et al [HQT93]. Cranston [Cra06$^+$] proved that the chromatic number of the square of a line graph of a graph of maximum degree $\Delta=4$ is at most 22, improving on the bound of 23, obtained by Horak [Hor90]. Note that this does not quite match the conjectured bound of 20. For larger Δ, Molloy and Reed [MR97] showed that there is an $\varepsilon>0$ such that the chromatic number of the square of the line graph of G is at most $(2-\varepsilon)\Delta(G)^2$.

It is not even known if the clique number of the square of the line graph of G is at most $\frac{5}{4}\Delta(G)^2$, although Chung et al. [CGTT90] did prove that a graph G whose line graph is a clique has at most $\frac{5}{4}\Delta(G)^2$ edges.

Inspired by the above conjecture, Faudree et al. [FGST90] proved that for bipartite G, the clique number of the line graph of G is at most Δ^2. $K_{\Delta,\Delta}$ shows that this bound is tight. They conjectured the same bound holds for the chromatic number ([FGST89] see also [BQ93]).

If every edge of G is in $\frac{3}{4}\Delta(G)^2$ cycles of length four, then the square of the line graph of G has maximum degree less than $\frac{5}{4}\Delta(G)^2 - 1$, so the result follows from Brooks' Theorem. Mahdian [Mah00] proved that if G has no C_4 then the square of its line graph has chromatic index $o(\Delta(G)^2)$. These two complementary results provide strong evidence that the conjecture holds, at least asymptotically. We refer the reader to Mahdian's M.Sc. thesis for a fuller discussion of this conjecture, including the origin of the use of the term *strong edge colouring* for a partition into induced matchings, and *strong chromatic index* for the chromatic number of the square of the line graph.

References

[And92] L. D. Andersen, The strong chromatic index of a cubic graph is at most 10, *Discrete Mathematics* **108** (1992), 231–252.

[BQ93] R. A. Brualdi and J. J. Quinn Massey, Incidence and strong edge colourings of graphs, *Discrete Mathematics* **122** (1993), 51–58.

[CGTT90] F. R. K. Chung, A. Gyárfás, W. T. Trotter, and Z. Tuza, The maximum number of edges in $2K_2$-free graphs of bounded degree, *Discrete Mathematics* **81**(1990), 129–135.

[Cra06+] D. Cranston, A strong edge coloring of graphs of maximum degree 4 using 22 colours, submitted (see `www.math.uiuc.edu/~cranston`).

[EN89] P. Erdős and J. Nešetřil, Problem, pp. 162–163 in G. Halász and V. T. Sós (eds.), *Irregularities of Partitions*, 1989.

[FGST89] R. J. Faudree, A. Gyárfás, R. H. Schelp, and Z. Tuza, Induced Matchings in Bipartite Graphs, *Discrete Mathematics* **78** (1989), 83–87.

[FGST90] R. J. Faudree, A. Gyárfás, R. H. Schelp, and Z. Tuza, The strong chromatic index of graphs, *Ars Combinatoria* **29B** (1990), 205–211.

[Hor90] P. Horak, The strong chromatic index of graphs of maximum degree four, pp. 399–403 in R. Bodendeik (Ed.) *Contemporary Methods in Graph Theory*, 1990.

[HQT93] P. Horak, H. Qing, and W. T. Trotter, Induced matchings in cubic graphs, *Journal of Graph Theory* **17** (1993), 151–160.

[Mah00] M. Mahdian, The strong chromatic index of C_4-free graphs, *Random Structures and Algorithms* **17**, pp. 357–375, 2000.

[MR97] M. Molloy and B. Reed, A bound on the strong chromatic index of a graph, *Journal of Combinatorial Theory (B)* **69** (1997), 103–109.

3 A Ramsey-type Problem on the Integers

Let $k \geq 3$ be fixed. We ask if there exist a $\delta > 0$ and a set $A \subseteq \mathbb{N}$ with the following properties:

(i) for every integer $\ell \geq 2$ every finite partition $A_1 \dot\cup A_2 \dot\cup \cdots \dot\cup A_\ell = A$ yields one partition class containing a k-AP, i.e., an arithmetic progression of length k and

(ii) every finite subset $A' \subseteq A$ contains a dense subset $A'' \subseteq A'$, $|A''| \geq \delta |A'|$ containing no k-AP.

Paul Erdős, Jaroslav Nešetřil, and Vojtěch Rödl in [ENR90].

> *"It was in the summer of 1983, when Paul Erdős came to Prague and, as usual, described many new conjectures that he first introduced Jarik and me to the problem of Pisier. We were initially optimistic as we thought our experience with ramsey type constructions might yield some results. However, after several failed attempts we began to suspect that Pisier's problem was beyond the scope of our abilities. Still, despite our disapointment, we found the problem very compelling and so, together with Paul Erdős, we considered several alternate versions of the original conjecture. The question presented here is one of those variations that, like the original problem, resisted all our efforts."*
>
> *Vojtěch Rödl*

This problem is motivated by the well known theorems of van der Waerden [Wae27] and Szemerédi [Sze75]. The theorem of van der Waerden [Wae27] is one of the earliest results in Ramsey theory. It asserts that every finite partition of the integers yields one partition class containing an arithmetic progression of any fixed length. More precisely, for positive integers k and ℓ, we say a set of integers A has the *van-der-Waerden-property* $\mathbf{vdW}(k, \ell)$ if for every partition $A_1 \dot\cup A_2 \dot\cup \cdots \dot\cup A_\ell = A$ there is some i $(1 \leq i \leq \ell)$ such that A_i contains a k-AP, i.e., an arithmetic progression of length k. The theorem of van der Waerden can then be stated as follows.

Theorem 3.1 (van der Waerden (1927)). *For all integers $k \geq 3$ and $\ell \geq 2$ there exist $n_{\mathrm{vdW}} = n_{\mathrm{vdW}}(k, \ell)$ such that for every $n \geq n_{\mathrm{vdW}}$ the set $[n] = \{1, 2, \ldots, n\}$ has $\mathbf{vdW}(k, \ell)$.*

Solving a longstanding standing conjecture of Erdős and Turán [ET36] Szemerédi proved the following famous generalization of Theorem 3.1, which stimulated a lot of research and today several proofs using tools from quite diverse areas of mathematics are known [Fur77, Gow06$^+$, Gow01, NRS06, RS04, Tao06$^+$b] (see also [Tao06$^+$a] for a survey of those proofs).

Theorem 3.2 (Szemerédi (1975)). *For every integer $k \geq 3$ and $\delta > 0$ there exist $n_{\mathrm{Sz}} = n_{\mathrm{Sz}}(k, \delta)$ such that for every $n \geq n_{\mathrm{Sz}}$ every subset $A \subseteq [n]$ with $|A| \geq \delta n$ contains a k-AP.*

Similarly as above we say a finite set of integers A has the *Szemerédi-property* $\mathbf{Sz}(k,\delta)$ if every subset $A' \subseteq A$ with $|A'| \geq \delta|A|$ contains a k-AP. Then Szemerédi's theorem asserts that every sufficiently large subset of the first n integers has $\mathbf{Sz}(k,\delta)$. Moreover, since Theorem 3.2 implies Theorem 3.1, it implies, e.g., that every sufficiently large arithmetic progression A displays both properties $\mathbf{vdW}(k,\ell)$ and $\mathbf{Sz}(k,\delta)$ and one may wonder if all sets of integers admitting the van-der-Waerden-property may have the Szemerédi-property as well. That would be somewhat surprising and a proof of such a statement would give a new proof of Szemerédi's theorem. Erdős, Nešetřil, and Rödl [ENR06+, ENR90] conjectured that this is not true. In other words, they conjectured that for fixed $k \geq 3$ there exist $\delta > 0$ and a set $A \subset \mathbb{N}$ which, on one hand, has the van-der-Waerden-property $\mathbf{vdW}(k,\ell)$ for every ℓ, but, on the other hand, no finite subset $A' \subseteq A$ has the Szemerédi-property $\mathbf{Sz}(k,\delta)$. For the case $k = 3$ a related question (motivated by this problem) was considered by Davenport, Hindman, and Strauss [DHS02].

The problem was also motivated by "negative" results concerning problems related to the well known problem of Pisier (see Problem 3.3 below). Suppose some family $\mathscr{I}$ of subsets of the integers is given. We call the elements $I \in \mathscr{I}$ *independent* sets. For an integer $k \geq 3$ let $\mathscr{I}_k = \{I \subseteq \mathbb{N}\colon I \text{ contains no } k\text{-AP}\}$. Then showing that no such set A with properties (i) and (ii) in the statement of the problem exists means to prove the following. *For every $\delta > 0$ and $A \subseteq \mathbb{N}$ there exist ℓ such that if every finite subset $A' \subseteq A$ contains an independent set $A'' \in \mathscr{I}_k$ of size $|A''| \geq \delta|A'|$, then $A = A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_\ell$ can be partitioned into ℓ independent sets, i.e., $A_i \in \mathscr{I}_k$ for every $i = 1, 2, \ldots, \ell$.* This formulation is formally related to Pisier's problem. To state this problem we say a set $I \subset \mathbb{N}$ is independent if all finite sums of I are distinct, i.e., for all finite, distinct subsets I_1, $I_2 \subseteq I$

$$\sum_{x \in I_1} x \neq \sum_{x \in I_2} x$$

and let $\mathscr{I}_P$ be the collection of all those sets. In [Pis83] Pisier asked whether the following is true.

Problem 3.3 (Pisier (1983)). For every $\delta > 0$ and $A \subseteq \mathbb{N}$ there exist ℓ such that if every finite subset $A' \subseteq A$ contains an independent set $A'' \in \mathscr{I}_P$ of size $|A''| \geq \delta|A'|$, then $A = A_1 \dot\cup A_2 \dot\cup \ldots \dot\cup A_\ell$ can be partitioned into ℓ independent sets, i.e., $A_i \in \mathscr{I}_P$ for every $i = 1, 2, \ldots, \ell$.

The affirmative answer of Problem 3.3 would give an arithmetic characterization of *Sidon sets* in terms of this condition.

As pointed out in [ENR06+] there are only very few non-trivial notions of independent families known, for which the Pisier-type problem was solved in the affirmative way. In [ENR06+] a few "negative" examples were shown, i.e., results which are formally similar to the problem.

References

[DHS02] D. Davenport, N. Hindman, and D. Strauss, Triangle free sets and arithmetic progressions – two Pisier type problems, *Electronic J. Combinatorics* **9** (2002), no. #R22, 1–19.

[ET36] P. Erdős and P. Turán, On some sequences of integers, *J. London. Math. Soc.* **11** (1936), 261–264.

[ENR06$^+$] P. Erdős, J. Nešetřil, and V. Rödl, On colorings and independent sets (Pisier type theorems), submitted.

[ENR90] P. Erdős, J. Nešetřil, and V. Rödl, On Pisier type problems and results (combinatorial applications to number theory), *Mathematics of Ramsey theory, Algorithms Combin.*, vol. 5, Springer, Berlin, 1990, pp. 214–231.

[Fur77] H. Furstenberg, Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, *J. Analyse Math.* **31** (1977), 204–256.

[Gow06$^+$] W. T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, submitted.

[Gow01] W. T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.* **11** (2001), no. 3, 465–588.

[NRS06] B. Nagle, V. Rödl, and M. Schacht, The counting lemma for regular k-uniform hypergraphs, *Random Structures Algorithms* **28** (2006), no. 2, 113–179.

[Pis83] G. Pisier, Arithmetic characterizations of Sidon sets, *Bull. Amer. Math. Soc. (N.S.)* **8** (1983), no. 1, 87–89.

[RS04] V. Rödl and J. Skokan, Regularity lemma for k-uniform hypergraphs, *Random Structures Algorithms* **25** (2004), no. 1, 1–42.

[Sze75] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.* **27** (1975), 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[Tao06$^+$a] T. Tao, The dichotomy between structure and randomness, arithmetic progressions, and the primes, submitted.

[Tao06$^+$b] T. Tao, A quantitative ergodic theory proof of Szemerédi's theorem, submitted.

[Wae27] B. L. van der Waerden, Beweis einer Baudetschen Vermutung, *Nieuw Arch. Wisk.* **15** (1927), 212–216, German.

4 The Pentagon Problem

Let G be a 3-regular graph that contains no cycle of length shorter than g. Is it true that for large enough g there is a homomorphism from G to C_5? Explicitly, is there a vertex coloring of G by $\{1, 2, 3, 4, 5\}$, such that colors of adjacent vertices differ by 1 modulo 5?

Jaroslav Nešetřil in [Neš99].

Apart from being published in [Neš99], this question was asked by Nešetřil at numerous problem sessions. By Brook's theorem any triangle-free cubic (i.e.

3-regular) graph is 3-colorable. Does a stronger assumption on girth of the graph imply existence of a more restricted coloring? (The girth of a graph G is the minimum length of a cycle in G.)

This problem is motivated by complexity considerations [GHN00] and also by exploration of density of the homomorphism order: We write $G \prec_h H$ if there is a homomorphism from G to H but not from H to G. It is known that whenever $G \prec_h H$ holds and H is not bipartite then there is a graph K satisfying $G \prec_h K \prec_h H$. In other words, the order $\prec_h$ is dense (if we do not consider edgeless graphs). A negative solution to the Pentagon problem would have the following density consequence: for each cubic graph H for which $C_5 \prec_h H$ holds, there exists a cubic graph K satisfying $C_5 \prec_h K \prec_h H$ (see [Neš99]).

If we replaced C_5 in the statement of the problem by a longer odd cycle, we would get a stronger statement. It is known that no such strenghthening is true. This was proved by Kostochka, Nešetřil, and Smolíková [KNS98] for C_{11} (hence for all C_l with $l \geq 11$), by Wanles and Wormald [WW01] for C_9, and recently by Hatami [Hat05] for C_7. Each of these results uses probabilistic arguments (random regular graphs), no constructive proof is known.

Häggkvist and Hell [HH93] proved that for every integer g there is a graph U_g with odd girth at least g (that is, U_g does not contain odd cycle of length less than g) such that every cubic graph of odd girth at least g maps homomorphically to U_g. Here, the graph U_g may have large degrees. This leads to a weaker version of the Pentagon problem: Is it true that for every k there exists a *cubic* graph H_k of girth k and an integer g such that every cubic graph of girth at least g maps homomorphically to H_k? A particular question in this direction: does a high-girth cubic graph map to the Petersen graph?

As an approach to this, we mention a result of DeVos and Šámal [DŠ06$^+$]: a cubic graph of girth at least 17 admits a homomorphism to the Clebsch graph. In context of the Pentagon problem, the following reformulation is particularly appealing: If G is a cubic graph of girth at least 17, then there is a cut-continuous mapping from G to C_5; that is, there is a mapping $f : E(G) \to E(C_5)$ such that for any cut $X \subseteq E(C_5)$ the preimage $f^{-1}(X)$ is a cut. (Here by *cut* we mean the edge-set of a spanning bipartite subgraph. A more thorough exposition of cut-continuous mappings can be found in [DNR02].)

References

[DNR02] M. DeVos, J. Nešetřil, and A. Raspaud, On flow and tension-continuous maps, *KAM-DIMATIA Series* **567** (2002).

[DŠ06$^+$] M. DeVos and R. Šámal, High girth cubic graphs map to the Clebsch graph, submitted.

[GHN00] A. Galluccio, P. Hell, and J. Nešetřil, The complexity of H-colouring of bounded degree graphs, *Discrete Math.* **222** (2000), no. 1–3, 101–109.

[HH93] R. Häggkvist and P. Hell, Universality of A-mote graphs, *European J. Combin.* **14** (1993), no. 1, 23–27.

[Hat05] H. Hatami, Random cubic graphs are not homomorphic to the cycle of size 7, *J. Combin. Theory Ser. B* **93** (2005), no. 2, 319–325.

[KNS98] A.V. Kostochka, J. Nešetřil, and P. Smolíková, Colorings and homomorphisms of degenerate and bounded degree graphs, *Discrete Math.* **233** (2001), no. 1–3, 257–276, Fifth Czech-Slovak International Symposium on Combinatorics, Graph Theory, Algorithms and Applications, (Prague, 1998).

[Neš99] J. Nešetřil, Aspects of structural combinatorics (graph homomorphisms and their use), *Taiwanese J. Math.* **3** (1999), no. 4, 381–423.

[WW01] I.M. Wanless and N.C. Wormald, Regular graphs with no homomorphisms onto cycles, *J. Combin. Theory Ser. B* **82** (2001), no. 1, 155–160.

5 Critical Graphs

Does every large k-critical graph contain a large $(k-1)$-critical subgraph?

Jaroslav Nešetřil and Vojtěch Rödl, International Colloquium on Finite and Infinite Sets, Keszthely, 1973.

> *"In 1973 Paul Erdős' 60th birthday was celebrated by the International Colloquium on Finite and Infinite Sets in Keszthely, Hungary. During the conference the participants had a memorable excursion by boat on Lake Balaton, with Erdős conducting a problem session onboard and the whole crowd visiting a vineyard on the northern coast. At the boat I met two young Czechoslovaks, Jaroslav Nešetřil and Vojtěch Rödl. They had asked Erdős whether every large k-critical graph always contains a large $(k-1)$-critical subgraph. Erdős obviously liked the problem, and knowing my interest in critical graphs [Toft70] he then got us in contact."*
>
> *Bjarne Toft*

A k-chromatic graph is k-critical if all proper subgraphs are $(k-1)$-colourable. For $k = 1$, 2 and 3 the k-critical graphs are the complete 1-graph K_1, the complete 2-graph K_2 and the odd cycles, respectively. For $k = 4$ the class of k-critical graphs is already very complicated. They are the forbidden subgraphs for 3-colourability, and it is an NP-complete problem type to decide about 3-colourability, as is well known. Thus for $k = 3$ the answer to the question in the title is obviously NO since 2-critical graphs have only two vertices and odd cycles may be large. However for $k = 4$ the situation is less clear. It turned out to be not so difficult to see that the answer for $k = 4$ is YES. However, for values of $k \geq 5$ the question is still unsettled.

The case $k = 4$

Does every large 4-critical graph contain a large odd cycle? Or more general: does every large 4-critical graph contain a large cycle? The answer to this second question was first given by Kelly and Kelly [KK54]. Let $L(n)$ denote the minimum taken over all 4-critical graphs G on n vertices of the maximum length of a cycle in G (this is called the circumference of G). Kelly and Kelly proved that indeed $L(n) \to \infty$ for $n \to \infty$. How fast does $L(n)$ tend to infinity? After subsequent improvements by Dirac [Dir55] and Reid [Reid57], Gallai [Gal63] obtained the so far best upper bound, namely that there is a constant c such that $L(n) < c \log n$. This means that the growth of the length of longest cycles in 4-critical graphs may be slow. It is seemingly still not known if this is best possible. The best lower bound is of order of magnitude $\sqrt{\log n}$, due to Alon, Krivelevich and Seymour [AKS00]. A large 4-critical graph therefore contains a long cycle. Since it is 2-connected and contains odd cycles, it is an easy exercise to show that it also must contain a long odd cycle. Thus the question of Nešetřil and Rödl has answer YES for $k = 4$. The case $k = 4$ was solved in a different manner by Voss [Voss77, Voss91]. He based his affirmative solution on the theory of bridges with respect to cycles in graphs.

The cases $k \geq 5$

Toft [Toft74] characterized the class of k-critical graphs in terms of the behaviour of the $(k-1)$-critical subgraphs they contain. One easy observation is that the $(k-1)$-critical subgraphs together cover the whole k-critical graph. In other words: any edge of a k-critical graph is contained in a $(k-1)$-critical subgraph. More generally, given two edges e_1 and e_2 of a k-critical graph, there is a $(k-1)$-critical subgraph containing e_1, but not e_2. The proof is simple, yet this seems to be useful. For example it follows easily from this that a k-critical graph is $(k-1)$-edge-connected, a result first obtained in a more complicated way by Dirac [Dir53]. Another consequence of the 'distinguishing property' of $(k-1)$-critical graphs was obtained by Stiebitz [Sti87]. He proved that if all $(k-1)$-critical subgraphs of a k-critical graph G are smallest possible, i.e. they are all complete $(k-1)$-graphs, then G is also smallest possible, i.e. G is the complete k-graph. This is related to the problem of Nešetřil and Rödl, giving an upper bound for the size of a k-critical graph in terms of its $(k-1)$-critical subgraphs, in a very special case. This problem was first thought of by Nešetřil and Toft during one of their later encounters, when they together visited G. A. Dirac at Aarhus in the mid 1970'ies. This special case, when all the $(k-1)$-critical subgraphs are complete, has the flavor of perfect graph theory, but is much, much easier to deal with (the main difference is that here we deal with all subgraphs, not just the induced ones). In connection with these results, the following is an interesting question: Given two arbitrary edges e_1 and e_2 of a k-critical graph with $k \geq 5$ is there a $(k-1)$-critical subgraph containing both e_1 and e_2? The answer is not known, even when the two edges e_1 and e_2 form a path of length 2. There seems to be no easy proof—this indicates that there may well be counterexamples. An example of a k-critical graph G, $k \geq 5$, without any $(k-1)$-critical subgraph containing two given edges e_1 and e_2 would be extremely interesting. Based on such a G one would be able to get a negative answer to the question of Nešetřil and Rödl, using copies of G and Hajós' construction [Haj61]:

Hajós' construction

Let G_1 and G_2 be disjoint graphs with edges x_1y_1 and x_2y_2 respectively. Remove x_1y_1 from G_1 and x_2y_2 from G_2, identify x_1 and x_2 to one new vertex x and join y_1 and y_2 by a new edge. Use this construction recursively on q disjoint copies G_1, G_2, …, G_q of the above G, with edges e_1 and e_2, removing edge e_2 from the copy G_i and edge e_1 from the copy G_{i+1}, $i = 1$, 2, …, $q-1$, identifying two endvertices from the removed edges and joining the two other ends by a new edge. The obtained k-critical graph H is large if q is large, yet any $(k-1)$-critical subgraph of H must be contained within two consecutive copies of G and hence be small (for $k \geq 5$).

Remarks

We saw in the previous sections that an example of a k-critical graph G, $k \geq 5$, containing edges e_1 and e_2 without any $(k-1)$-critical subgraph containing e_1 and e_2 would give the answer NO to the question of Nešetřil and Rödl. We know however that the answer is YES for $k = 4$. The case $k = 4$ behaves differently, and in fact for any 4-critical graph G and any path P of length 2 in G, there is an odd cycle in G containing P. This statement follows from an argument of Dirac [Dir64] and was also obtained by Wessel [Wes81]. The above potential counterexamples to the question of Nešetřil and Rödl have separating sets of size 2. It seems likely that such counterexamples exist. However most probably no counterexample is of connectivity at least 3 (or high enough). Is it possible (easy?) to prove that any large k-critical graph of connectivity at least 3 (or at least $c(k)$) contains a large $(k-1)$-critical subgraph? If we instead of just subgraphs ask for induced subgraphs, then it is not clear what to expect and what is known and what is not. The best way to look at this is perhaps to consider vertex-k-critical graphs, i.e. graphs G that are k-chromatic and $G - x$ is $(k-1)$-colourable for all vertices x in G. If all induced vertex-i-critical subgraphs of a vertex-k-critical graph G, $k \geq 4$, are smallest possible, i.e. complete i-graphs for all $i < k$, then G is small, more precisely G is either the complete k-graph or G is an odd cycle complement (with $2k-1$ vertices). As observed first by Wessel ([Wes77], see also [Toft85]) this is an equivalent statement to the very deep strong perfect graph conjecture, recently proved by Chudnovsky, Robertson, Seymour and Thomas [CRST06$^+$].

References

[AKS00] N. Alon, M. Krivelevich and P. Seymour, Long cycles in critical graphs, *J. Graph Theory* 35 (2000), 193–196.

[CRST06+] M. Chudnovsky, N. Robertson, P. Seymour and R. Thomas, The strong perfect graph theorem, *Annals of Mathematics*, to appear.

[Dir53] G. A. Dirac, The structure of k-chromatic graphs, *Fund. Math.* 40 (1953), 42–55.

[Dir55] G. A. Dirac, Circuits in critical graphs, *Monatsh. Math.* 59 (1955), 178–187.

[Dir64] G. A. Dirac, On the structure of 5- and 6-chromatic abstract graphs, *J. Reine Angew. Math.* 214/215 (1964), 43–52.

[Gal63] T. Gallai, Kritische Graphen I, *Publ. Math. Inst. Hungar. Acad. Sci.* 8 (1963), 165–192.

[Haj61] G. Hajós, Über eine Konstruktion nicht n-färbbarer Graphen, *Wiss. Z. Martin-Luther-Univ. Halle-Wittenberg Math.-Natur. Reihe* 10 (1961), 116–117.

[KK54] J. B. Kelly and L. M. Kelly, Paths and circuits in critical graphs, *Amer. J. Math.* 76 (1954), 786–792.

[Reid57] R. C. Reid, Maximal circuits in critical graphs, *J. London Math. Soc.* 32 (1957), 456–462.

[Sti87] M. Stiebitz, Subgraphs of colour-critical graphs, *Combinatorica* 7 (1987), 303–312.

[Toft70] B. Toft, *Some contributions to the theory of colour-critical graphs*, Ph.D.-thesis University of London 1970, published by Aarhus University in Various Publication Series 14 (1970).

[Toft74] B. Toft, On critical subgraphs of colour-critical graphs, *Discrete Math.* 7 (1974), 377–392.

[Toft85] B. Toft, Some problems and results related to subgraphs of colour critical graphs, in: R. Bodendiek, H. Schumacher and G. Walther (editors), *Graphen in Forschung und Unterricht, Festschrift K. Wagner*, Barbara Franzbecker Verlag 1985, 178–186.

[Voss77] H.-J. Voss, Graphs with prescribed maximal subgraphs and critical chromatic graphs, *Comment. Math. Univ. Carolinae* 18 (1977), 129–142.

[Voss91] H.-J. Voss, *Cycles and bridges in graphs*, Deutscher Verlag der Wissenschaften, Kluwer Academic Publishers, 1991.

[Wes77] W. Wessel, Some colour-critical equivalents of the strong perfect graph conjecture, in: Proc. Int. Koll. Graphentheorie und deren Anwendungen, Oberhof DDR, Matematisches Gesellschaft der DDR 1977, 300–309.

[Wes81] W. Wessel, Critical lines, critical graphs and odd cycles, Technical report, Akademie der Wissenschaften der DDR, Institut für Mathematik 1981.

6 The CLIQUE Problem in Geometric Intersection Graphs

Determine the computational complexity of the CLIQUE problem restricted to intersection graphs of straight line segments in the plane.

Jan Kratochvíl and Jaroslav Nešetřil in [KN90].

> *"This recollection illustrates Jarik Nešetřil's gentle understanding of students' feelings, as well as his excellent instinct in finding rewarding problems. Back in the beginning of the 1980's a group of undergraduate students of Charles University in Prague was discussing an urgent matter of selecting a research seminar. None of the officially offered ones was a winning favorite, and we had just three days to file our decision at the registrar. The discussion was held during a regular Graph Theory lecture of Jarik. Instead of expelling us from the class for disturbing, he quickly got himself involved in the discusion and on the spot offered to create a seminar especially for us. Who could resist such a generous offer? He also suggested a problem to work on. In the following year or two we learned a lot about research while working on the problem of characterizing string graphs. Though we did not manage to characterize these graphs, many side results led to a conference presentation and an undergraduate publication. I revived the problem for myself about 5 years later when I finally proved NP-hardness of the recognition problem. And when presenting this negative solution to Jarik, we started discussing the complexity of optimization problems in restricted classes of graphs and the CLIQUE question was born."*
>
> *Jan Kratochvíl*

An abstract graph G is a $\mathcal{K}$-intersection graph, for some class $\mathcal{K}$ of sets, if the vertices of G can be represented by sets in $\mathcal{K}$ such that two vertices in G are adjacent iff the corresponding sets have a nonempty intersection.

Intersection graphs for various classes of geometric objects (e.g., straight line segments, rectangles, or disks in the plane) have been studied extensively. On the one hand, they have numerous practical applications (for instance, in frequency assignment in cellular networks [Hale80, Mal97, AHKMS01], or in map labelling [AKS97]). On the other hand, geometric intersection graphs provide a rich source of classes of graphs with interesting properties, and of challenging problems that lie at the interface between graph theory and geometry.

Deciding for graph G if it is an intersection graph of a certain kind, or computing a representation, is often computationally hard. For instance, this recognition problem is $\mathcal{NP}$-hard for intersection graphs of disks [HK01] and of segments [KM94], respectively. Furthermore, representations may require coordinates that are exponential in the size of the graph [KM94], so it is not clear if these problems even belong to the class $\mathcal{NP}$; only $\mathcal{PSPACE}$ membership is known [BK98, KM94].

A circle of questions naturally arising in the applications concern the complexity of classical hard problems, such as CLIQUE or INDEPENDENT SET, for intersection graphs. In some cases, many of such problems become tractable; for instance, CLIQUE is polynomially solvable for intersection graphs of equal-radius disks [CCJ90], or of segments with a bounded number of directions [KN90]. For both results, it is assumed that a suitable form of geometric representation is provided as part of the input, because recognition remains $\mathcal{NP}$-hard also under the additional assumptions.

Even when the problem remains hard, the geometric structure might lead to better approximation algorithms. For instance, for general graphs, it is hard to approximate the size of a maximum independent set within a factor of $n^{1-\varepsilon}$ [Hås99], for any fixed $\varepsilon > 0$. Exactly solving INDEPENDENT SET remains $\mathcal{NP}$-hard in intersection graphs of segments [KN90] (even if the segments are restricted to 2 directions) and of disks [CCJ90] (even if the disks all have the same radius). However, the problem can be approximated in polynomial time within a factor of roughly $\sqrt{n}$ for intersection graphs of segments [AM04], and within $(1 + \varepsilon)$, for any fixed $\varepsilon > 0$, in the case of disks [HMRRRS98, EJS05, Chan03]. For unit disks, even the assumption that a representation is provided can be avoided [NHK04].

Among the most tantalizing unsolved problems in this area are the complexity of the CLIQUE problem for intersection graphs of segments and of disks, respectively. For segments, the question was first posed by Kratochvíl and Nešetřil in 1990, while for disk graphs, it seems to be folklore. We remark that the above-mentioned algorithm for equal-radius disks breaks down as soon as two radii are allowed, while for the case of segments with a bounded number d of directions, the runtime of the algorithm depends exponentially on d. As for results in the opposite direction, every complement of a planar graph can be represented as an intersection graph of convex polygons in the plane [KK98]. It follows that CLIQUE is $\mathcal{NP}$-hard for such graphs, because INDEPENDENT SET is hard for planar graphs. The polygons used in the representation are of nonconstant complexity. There are results for two types of geometric objects of constant complexity. The first type are intersection graphs of angles [MP92], where an angle consist of one horizontal and one vertical segment sharing a common endpoint. If all the angles are "upper" ones, say, then CLIQUE is polynomially solvable, but if opposite angles are allowed, then the problem is $\mathcal{NP}$-hard. The second type are intersection graphs of ellipses [AW05]. For these, CLIQUE is $\mathcal{NP}$-hard. In fact, for sufficiently small $\delta > 0$, even approximation within $(1+\delta)$ is hard. Moreover, this continues to hold even if for all ellipses, the ratio between the two principal radii is required to be any given constant ρ, $1 < \rho < \infty$. However, the "limit cases" of circles ("$\rho = 1$") and of segments ("$\rho = \infty$"), respectively, remain open.

References

[AHKMS01] K. I. Aardal, S. P. M. van Hoesel, A. M. C. A. Kosterand, C. Mannino, and A. Sassano, Models and solution techniques for frequency assignment problems, Technical report, Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2001.

[AM04] P. K. Agarwal and N. H. Mustafa, Independent set of intersection graphs of convex objects in 2d, in *9th Scandinavian Workshop on Algorithm Theory (SWAT'04)*, pages 127–137, 2004.

[AKS97] P. K. Agarwal, M. van Kreveld, and S. Suri, *Label Placement by Maximum Independent Set in Rectangles*, pages 233–238, 1997.

[AW05] C. Ambühl and U. Wagner, The clique problem in intersection graphs of ellipses and triangles, *Theory Comput. Syst.*, 38(3):279–292, 2005.

[BK98] H. Breu and D. G. Kirkpatrick, Unit disk graph recognition is NP-hard. *Computational Geometry. Theory and Applications*, 9(1-2):3–24, 1998.

[Chan03] T. M. Chan, Polynomial-time approximation schemes for packing and piercing fat objects, *J. Algorithms*, 46(2):178–189, 2003.

[CCJ90] B. N. Clark, C. J. Colbourn, and D. S. Johnson, Unit disk graphs, *Discrete Mathematics*, 86:165–177, 1990.

[EJS05] T. Erlebach, K. Jansen, and E. Seidel, Polynomial-time approximation schemes for geometric intersection graphs, *SIAM J. Comput.*, 34(6):1302–1323, 2005.

[Hale80] W. K. Hale, Frequency assignment: Theory and applications, *Proceedings of the IEEE*, 68(12):1497–1514, December 1980.

[Hås99] J. Håstad, Clique is hard to approximate within $n^{1-\varepsilon}$, *Acta Math.*, 182(1):105–142, 1999.

[HK01] P. Hliněný and J. Kratochvíl, Representing graphs by disks and balls (a survey of recognition-complexity results), *Discrete Math.*, 229(1-3):101–124, 2001.

[HMRRRS98] H. B. Hunt, III, M. V. Marathe, V. Radhakrishnan, S. S. Ravi, D. J. Rosenkrantz, and R. E. Stearns, NC-approximation schemes for NP- and PSPACE-hard problems for geometric graphs, *J. Algorithms*, 26(2):238–274, 1998.

[KK98] J. Kratochvíl and A. Kuběna, On intersection representations of co-planar graphs, *Discrete Math.*, 178(1-3):251–255, 1998.

[KM94] J. Kratochvíl and J. Matoušek, Intersection graphs of segments, *J. Combin. Theory Ser. B*, 62(2):289–315, 1994.

[KN90] J. Kratochvíl and J. Nešetřil, Independent set and clique problems in intersection-defined classes of graphs, *Comment. Math. Univ. Carolin.*, 31(1):85–93, 1990.

[Mal97] E. Malesińska, *Graph-Theoretical Models for Frequency Assignment Problems*, PhD thesis, TU Berlin, 1997.

[MP92] M. Middendorf and F. Pfeiffer, The max clique problem in classes of string-graphs, *Discrete Math.*, 108(1-3):365–372, 1992.

[NHK04] T. Nieberg, J. Hurink, and W. Kern, A robust PTAS for maximum weight independent sets in unit disk graphs, In *Proceedings of the 30th International Workshop on Graph-Theoretic Concepts in Computer Science (WG'04)*, 2004.