Problems in Ramsey Theory, Probabilistic Combinatorics and Extremal Graph Theory

Bhargav Peruvemba Narayanan

Trinity College
and
Department of Pure Mathematics and Mathematical Statistics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

October 29, 2015

Declaration

This document is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. The results in Chapters 2 and 4 were obtained in collaboration with T. Kittipassorn and my contribution was about 50%. Chapter 5 is based on joint work with B. Bollobás and A. Raigorodskii and in this case, I did the majority of the work. Chapter 6 is based on joint work with J. Balogh and B. Bollobás and my contribution was about 60%. The results in Chapter 7 were obtained in collaboration with with P. Balister, B. Bollobás and J. Lee and my contribution was about 50%. Chapter 8 is based on joint work with B. Bollobás, T. Kittipassorn and A. Scott and my contribution was about 40%. Finally, Chapter 9 is based on joint work with P. Balister, S. Binski and B. Bollobás and my contribution was about 50%. No part of this dissertation has been submitted for any other qualification.

Bhargav Peruvemba Narayanan

Abstract

In this dissertation, we treat several problems in Ramsey theory, probabilistic combinatorics and extremal graph theory.

We begin with the Ramsey theoretic problem of finding exactly m-coloured graphs. For which natural numbers $m \in \mathbb{N}$ are we guaranteed to find an m-coloured complete subgraph in any edge colouring of the complete graph on $\mathbb{N}$? We resolve this question completely and prove, answering a question of Stacey and Weidl [104], that whenever we colour $\mathbb{N}^{(2)}$ with infinitely many colours, we are guaranteed to find an $\binom{n}{2}$-coloured complete subgraph for each $n \in \mathbb{N}$. In addition, we also demonstrate that given a colouring of $\mathbb{N}^{(2)}$ with k colours, there are at least $\sqrt{2k}$ distinct values $m \in [k]$ for which an infinite m-coloured complete subgraph exists. Finally, we also prove that given a colouring of $\mathbb{N}^{(2)}$ with k colours and $m \in [k]$, we can always find an infinite $\hat{m}$-coloured complete subgraph for some $\hat{m} \in [k]$ such that $|m - \hat{m}| \leq \sqrt{m/2}$.

Next, we give some results in probabilistic combinatorics. First, we investigate the stability of the Erdős–Ko–Rado Theorem. For natural numbers $n, r \in \mathbb{N}$ with $n \geq r$, the Kneser graph $K(n, r)$ is the graph on the family of r-element subsets of $\{1, \ldots, n\}$ in which two sets are adjacent if and only if they are disjoint. Delete the edges of $K(n, r)$ with some probability, independently of each other: is the independence number of this random graph equal to the independence number of the Kneser graph itself? We shall answer this question affirmatively as long as r/n is bounded away from $1/2$, even when the probability of retaining an edge of the Kneser graph is quite small; we also prove a much more precise result when $r = o(n^{1/3})$. We then study a geometric bootstrap percolation model on the three dimensional grid $[n]^3$ called *line percolation*. In line percolation with infection parameter r, infection spreads from a subset $A \subset [n]^3$ of initially infected lattice points as follows: if there is an axis parallel line $\mathcal{L}$ with r or more infected lattice points on it, then every lattice point of $[n]^3$ on $\mathcal{L}$ gets infected and we repeat this until the infection can no longer spread. Our main result is the determination the critical density of initially infected points at which percolation (infection of the entire grid) becomes likely.

Finally, we present two results in extremal graph theory. First, we consider a graph partitioning problem. For a graph G, let $f(G)$ be the largest integer k such that there are two vertex-disjoint subgraphs of G, each on k vertices, inducing the same number of edges. We prove that $f(G) \geq n/2 - o(n)$ for every graph G on n vertices, settling a conjecture of Caro and Yuster [36]. Finally, we study the problem of cops and robbers on the grid where the robber is allowed to move faster than the cops. We prove that when the speed of the robber is a sufficiently large constant, the number of cops needed to catch the robber on an $n \times n$ grid is $\exp(\Omega(\log n / \log \log n))$.

Acknowledgements

I am very grateful to my supervisor, Béla Bollobás, for all that he has done for me over the last three years. I cannot thank him enough for his guidance and generosity, all the beautiful mathematics he has introduced me to, and the many fantastic research visits he has organised. I would also like to thank Béla and his wonderful wife Gabi for all their support in general, and the many enjoyable evenings full of American football in particular.

I am grateful to Ramarathnam 'Venkie' Venkatesan for introducing me to the world of mathematical research and for always believing in me; he has been a source of great support and excellent advice, mathematical and otherwise.

I would like to thank all my collaborators: Paul Balister, József Balogh, Scott Binski, Victor Falgas-Ravry, Teeradej Kittipassorn, Dániel Korándi, Jonathan Lee, Shoham Letzter, Andrei Raigorodskii, Julian Sahasrabudhe, Alex Scott and István Tomon. I would also like to thank Sean Eberhard, Josh Erde, Imre Leader and Mark Walters for many interesting discussions.

My thanks to Richard Johnson, Tomas Juškevičius, Sebastian Koch, Michał Przykucki, Stephen Smith and all the other members of Team Bollobás for the mini-seminars, the coffee time problem sessions and all the other fun things we have done together.

I am grateful to Trinity College for providing me with the financial support for my work. I should also like to thank the University of Memphis, the Alfréd Rényi Institute of Mathematics, Budapest, the Theory Group at Microsoft Research, Redmond, and the IMT Institute for Advanced Studies, Lucca for their hospitality.

I would like to thank Juhan Aru, Tom Gillespie, Vytautas Gruslys, Arès Méroueh, Gunnar Peng, Paul Smith, Moses Tanenbaum, Noam Zamir and my other friends in Cambridge for making my time here so memorable. My thanks also to Adhiraj Alai, Advait Alai, Neal Bushaw, Audrius Feigelovicius, Rajkishan Gunasekaran, Immanuel Thomas and all my other friends in far off places for their support.

Finally, I would like to thank my parents, Sarala and Raju, for everything. Without their love and encouragement, none of this would have been possible.

Contents

Chapter 1. Introduction 1

1. Ramsey theory 1
2. Probabilistic combinatorics 2
3. Extremal graph theory 4

Part 1. Ramsey theory 7

Chapter 2. A canonical Ramsey theorem for exactly m-coloured graphs 9

1. Introduction 9
2. Our results 9
3. Proof of the main theorem 13
4. Extensions and applications 20
5. Concluding remarks 24

Chapter 3. Infinite exactly m-coloured complete graphs 27

1. Introduction 27
2. Lower bounds 30
3. Upper bounds 35
4. Concluding remarks 36

Chapter 4. Approximations to infinite m-coloured complete hypergraphs 39

1. Introduction 39
2. Proofs of the main results 42
3. Concluding remarks 50

Part 2. Probabilistic combinatorics 51

Chapter 5. Transference for the Erdős–Ko–Rado theorem, I 53

1. Introduction 53
2. Preliminaries 57
3. Upper bound for the critical threshold 59
4. Lower bound for the critical threshold 67
5. Concluding remarks 68

Chapter 6. Transference for the Erdős–Ko–Rado theorem, II 71
1. Introduction 71
2. Preliminaries 72
3. The number of disjoint pairs 74
4. Proof of the main result 78
5. Avenues for improvement 81
6. Concluding remarks 87

Chapter 7. Line percolation 89
1. Introduction 89
2. Our results 91
3. Probabilistic preliminaries 95
4. Line percolation in two dimensions 96
5. The critical probability in three dimensions 104
6. Minimal percolating sets 113
7. Concluding remarks 114

Part 3. Extremal graph theory 115

Chapter 8. Disjoint induced subgraphs of the same order and size 117
1. Introduction 117
2. Preliminaries 119
3. Overview of our strategy 124
4. Proof of the main result 126
5. Concluding remarks 145

Chapter 9. Catching a fast robber on the grid 147

1. Introduction 147
2. Proof of the main result 149
3. Concluding remarks 158

Bibliography 159

CHAPTER 1

Introduction

This dissertation is divided into three parts. The first part covers some Ramsey theory, the second is devoted to questions from probabilistic combinatorics and the third part deals with some problems in extremal graph theory. In what follows, we briefly discuss the problems and the results presented in the subsequent chapters.

1. Ramsey theory

Ramsey theory is concerned with questions about whether one can always find homogeneous substructures in large discrete structures. These questions are typically phrased in terms of colourings; a homogeneous substructure in this case is usually a substructure on which the colouring is particularly simple to describe. The first part of the dissertation consists of three chapters which are concerned with the following question: given a large coloured structure and a natural number $m \in \mathbb{N}$, can one always find a substructure that is coloured with exactly m distinct colours?

Given an edge-coloured graph, call a subgraph m-coloured if the colouring attains exactly m different values in the subgraph. For a natural number $m > 1$, it is natural to ask if one is guaranteed to find an exactly m-coloured complete subgraph in every edge-colouring of the complete graph on $\mathbb{N}$ with sufficiently many colours. It is not clear that there are any natural numbers with this property; the injective colouring shows, for example, that there is no such guarantee unless $m = \binom{n}{2}$ for some $n \in \mathbb{N}$. In Chapter 2, which is joint work with T. Kittipassorn [76], we answer a question of Stacey and Weidl [104] and prove that whenever we colour $\mathbb{N}^{(2)}$ with infinitely many colours, we are guaranteed to find an $\binom{n}{2}$-coloured complete subgraph for each $n \in \mathbb{N}$. We need two key ideas to prove this theorem. The first is to ask a more general question, the answer to which is a canonical Ramsey theorem for m-coloured complete subgraphs that implies the result we would like to prove. The second is that to prove this canonical Ramsey theorem (which is about edge-coloured complete graphs), somewhat strangely, we need to suitably generalise the result to general edge-coloured graphs (as opposed to complete graphs).

In Chapter 3, we investigate the following question: given a colouring of $\mathbb{N}^{(2)}$ with k colours, for how many distinct values $m \in [k]$ are we guaranteed to find an infinite m-coloured complete subgraph? We show in [91] that we are guaranteed at least $\sqrt{2k}$ distinct values and that this bound is tight infinitely often. We prove this result by proving a more intricate (and stronger) structural result by induction.

In Chapter 4, which is joint work with T. Kittipassorn [75], we study the following related problem: given a colouring of $\mathbb{N}^{(2)}$ with k colours and $m \in [k]$, how close can we get to finding an infinite m-coloured complete subgraph? Our main result is that we can always find an infinite $\hat{m}$-coloured complete subgraph for some $\hat{m} \in [k]$ such that $|m - \hat{m}| \le \sqrt{m/2}$, and that this is best possible. In the process, we also resolve a conjecture of mine from [91]. We also study how one might generalise this result to uniform hypergraphs; we consider two natural variants one of which we prove to be true while showing the other to be false.

2. Probabilistic combinatorics

The first two chapters of the second part of this dissertation are concerned with a 'sparse-random' analogue of the Erdős–Ko–Rado theorem. The Erdős–Ko–Rado Theorem is a central (and very simple) result in extremal set theory which tells us how large uniform intersecting families can be. One possible formulation of the Erdős–Ko–Rado theorem is the following: if $n \ge 2r$, then the size of the largest independent set of the graph $K(n, r)$ is $\binom{n-1}{r-1}$, where $K(n, r)$

is the *Kneser graph* with parameters $n, r \in \mathbb{N}$; the vertex set of this graph is the family of r-element subsets of $\{1, \dots, n\}$, and two r-sets are adjacent in $K(n, r)$ if and only if they are disjoint.

Let us delete the edges of the Kneser graph with some probability, independently of each other; is the independence number of this random graph equal to the independence number of the Kneser graph itself? Chapters 5 and 6 are both concerned with this question.

In Chapter 5, which is based on joint work with B. Bollobás and A. Raigorodskii [30], we answer the question affirmatively when $r = o(n^{1/3})$ and also determine the precise critical threshold at which the answer becomes negative; it turns out that a random analogue of the Erdős–Ko–Rado theorem continues to hold even after we have deleted practically all the edges of the Kneser graph.

Chapter 6, which is based on joint work with J. Balogh and B. Bollobás [17], contains a more substantial result: we show the answer to be in the affirmative as long as r/n is bounded away from $1/2$. To prove this result, we make use of a variety of tools. For example, we give some new estimates for the number of disjoint pairs in a family in terms of its distance from an intersecting family. We also briefly describe how ideas from the theory of graph containers can help sharpen these results.

In the third chapter of this part, we study a geometric bootstrap percolation model on the d-dimensional grid $[n]^d$. In this model, *line percolation*, with infection parameter r, infection spreads from a subset $A \subset [n]^d$ of initially infected lattice points as follows: if there is an axis parallel line $\mathcal{L}$ with r or more infected lattice points on it, then every lattice point of $[n]^d$ on $\mathcal{L}$ gets infected and we repeat this until the infection can no longer spread. The elements of the set A are usually chosen independently, with some density p, and the main question is to determine $p_c(n, r, d)$, the density at which percolation (infection of the entire grid) becomes likely. As is often the case with bootstrap percolation models, analysing the process in three dimensions turns out be significantly more challenging than the corresponding problem in two dimensions. Our main result in Chapter 7, which is based on joint work with P. Balister, B. Bollobás and J. Lee [11], is the determination of the critical probability in three dimensions up to multiplicative constants. The crux of the proof is in showing that long-range interactions are not very helpful in spreading the infection. This is done with a coupling argument where we run the process and wait for certain 'bad' events to occur; we then restart the process with the 'badness' built in from the start, but also impose additional constraints on how the infection might spread which makes this modified process easier to follow. We also determine the size of minimal percolating sets using the polynomial method.

3. Extremal graph theory

The first of the two chapters of this part, Chapter 8, is based on joint work with B. Bollobás, T. Kittipassorn and A. Scott [28], and is concerned with a graph partitioning problem. We would like to find two vertex-disjoint induced subgraphs of a given graph of the same order and size; how large can we guarantee these subgraphs to be? We answer this question and settle a conjecture of Caro and Yuster by showing that any n-vertex graph contains two very large, indeed of order $n/2 - o(n)$, disjoint induced subgraphs of the same order and size. The main idea is that for each graph, there is a (small) probability such that if we delete vertices from the graph independently with this probability, the resulting graph has many useful 'gadgets' (pairs of vertices that have suitable degree differences which we can use to find a suitable partition of this graph). However, since the argument has to cover all graphs, there is no simple description of this probability and we need to distinguish a few cases and tailor the argument suitably to fit each case.

The second chapter of this part, Chapter 9, is concerned with the game of cops and robbers. It is well known that in the traditional variant of the game of cops and robbers on an $n \times n$ grid, two cops are necessary and sufficient to catch the robber. In Chapter 9, which is based on joint work with P. Balister,

S. Binski and B. Bollobás [10], we ask how many cops are needed to catch the robber if the robber is allowed to move faster than the cops. We show that if the robber's speed is greater than some large constant, then at least $n^{1/\log\log n}$ cops are needed to catch the robber on an $n \times n$ grid. While this seems to be very far from the truth (indeed, we think the number of cops needed to catch a fast robber should be almost, if not actually, linear in n), the proof strategy might be of independent interest: we use a dynamic variant of the density-based strategy used by Bollobás and Leader to resolve Conway's Angel and Devil problem in three dimensions.

Part 1

Ramsey theory

CHAPTER 2

A canonical Ramsey theorem for exactly m-coloured graphs

Joint work with Teeradej Kittipassorn.

1. Introduction

A classical result of Ramsey [96] says that when the edges of a complete graph on a countably infinite vertex set are finitely coloured, one can always find a complete infinite subgraph all of whose edges have the same colour.

Ramsey's theorem has since been generalised in many ways; most of these generalisations are concerned with finding other monochromatic structures. For a survey of many of these generalisations, see the book of Graham, Rothschild and Spencer [63]. Ramsey theory has witnessed many developments over the last fifty years and continues to be an area of active research today; see, for instance, [106, 39, 69, 82, 22].

Alternatively, anti-Ramsey theory, which originates in a paper of Erdős, Simonovits and Sós [51], is concerned with finding large 'rainbow coloured' or 'totally multicoloured' structures. Between these two ends of the spectrum, one could consider the question of finding structures which are coloured with exactly m different colours as was first done by Erickson [52]; it is this line of enquiry that we pursue here.

2. Our results

Our notation is standard. We write $[n]$ for $\{1, ..., n\}$, the set of the first n natural numbers. For a set X, we write $X^{(2)}$ for the family of all two element subsets of X; equivalently, $X^{(2)}$ is the complete graph on the vertex set X. We denote a surjective map f from a set X to another set Y by $f : X \twoheadrightarrow Y$. By a *colouring* of a graph, we mean a colouring of the edges of the graph.

Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$ be a surjective colouring of the edges of the complete graph on $\mathbb{N}$ with an arbitrary set of colours $\mathcal{C}$. If the set of colours $\mathcal{C}$ is infinite, we say that Δ is an *infinite-colouring* and if $\mathcal{C}$ is finite, we say that Δ is a *k-colouring* if $|\mathcal{C}| = k$. We say that a subset X of $\mathbb{N}$ is (*exactly*) *m-coloured* if $\Delta(X^{(2)})$, the set of values attained by Δ on the edges with both endpoints in X, has size exactly m. We write $\gamma_\Delta(X)$, or $\gamma(X)$ in short, for the size of the set $\Delta(X^{(2)})$; in other words, every set X is $\gamma(X)$-coloured.

This chapter is concerned with canonical Ramsey theory, a subject which originates in a classical paper of Erdős and Rado [47]. To give readers unfamiliar with this subject a flavour of the results in this area, we recall a basic canonical Ramsey theorem proved by Erdős and Rado. To state this result, it will be convenient to introduce some notation. We say that $X \subset \mathbb{N}$ is *rainbow coloured* if no two edges with both endpoints in X receive the same colour. Also, we say that $X \subset \mathbb{N}$ is *left coloured* if for $i, j, k, l \in X$ with $i < j$ and $k < l$, $\Delta(ij) = \Delta(kl)$ if and only if $i = k$, and the definition of *right coloured* is analogous; if X is left or right coloured, we say, in short, that X is *lexically coloured*. With these definitions in place, we can now state the canonical Ramsey theorem of Erdős and Rado [47].

Theorem 2.1. *For any colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$, there exists an infinite subset X of $\mathbb{N}$ such that either*

(1) X is 1-coloured, or

(2) X is rainbow coloured, or

(3) X is lexically coloured. □

Returning to the question at hand, our main aim in this chapter is to prove a canonical Ramsey theory for m-coloured graphs. For a colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$

of the complete graph on $\mathbb{N}$ with an arbitrary set of colours, we define the set

$$\mathcal{G}_\Delta = \{\gamma_\Delta(X) : X \subset \mathbb{N}\}.$$

Stacey and Weidl [104] considered the following question: are there natural numbers $m \in \mathbb{N}$ that are guaranteed to be elements of $\mathcal{G}_\Delta$ for every infinite-colouring Δ? It is clear that the natural number 1 trivially has this property since an edge is a 1-coloured complete graph; however, it is not at all clear why any natural number $m > 1$ should have this property. By considering a rainbow colouring of $\mathbb{N}$, we see that $m \in \mathbb{N}$ cannot have this property unless $m = \binom{n}{2}$ for some $n \geq 2$. On the other hand, Stacey and Weidl were able to show that $\binom{3}{2}$ is always an element of $\mathcal{G}_\Delta$ for every infinite-colouring Δ. But for $n \geq 4$, they were unable to decide whether or not there exists an infinite-colouring Δ such that $\binom{n}{2} \notin \mathcal{G}_\Delta$. In particular, they asked if all natural numbers of the form $\binom{n}{2}$ must be contained in $\mathcal{G}_\Delta$ for every infinite-colouring Δ.

Here, we shall consider a more general question: when is $\mathcal{G}_\Delta \neq \mathbb{N}$? As remarked above, for an injective colouring Δ, $\mathcal{G}_\Delta = \{\binom{n}{2} : n \geq 2\} \neq \mathbb{N}$. There is another infinite-colouring Δ for which $\mathcal{G}_\Delta \neq \mathbb{N}$ which is slightly less obvious. Given $X \subset \mathbb{N}$, if there is a vertex $v \in X$ such that $X \setminus \{v\}$ is 1-coloured and all the edges between v and $X \setminus \{v\}$ have distinct colours (which are also all different from the colour appearing in $X \setminus \{v\}$), then we say that X is *star coloured* (*with centre* v). It is easy to check (see Figure 1) that if $\mathbb{N}$ is star coloured by Δ, then $\mathcal{G}_\Delta = \mathbb{N} \setminus \{2\}$.

Our main result, stated below, is that the two colourings described above are, in a sense, the 'canonical' colourings for which $\mathcal{G}_\Delta \neq \mathbb{N}$.

Theorem 2.2. *For every infinite-colouring* $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathbb{N}$, *either*

(1) $\mathcal{G}_\Delta = \mathbb{N}$, *or*

(2) there exists an infinite rainbow coloured subset of $\mathbb{N}$, *or*

(3) there exists an infinite star coloured subset of $\mathbb{N}$.

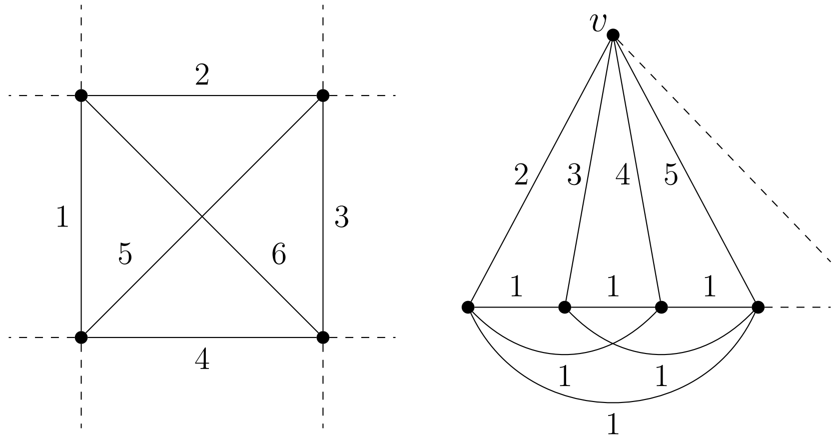

FIGURE 1. A rainbow colouring and a star colouring with centre v.

An immediate consequence of Theorem 2.2 is that the answer to the question posed by Stacey and Weidl is in the affirmative.

COROLLARY 2.3. *For every infinite-colouring* $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathbb{N}$, *and for every natural number* $n \geq 2$, $\binom{n}{2} \in \mathcal{G}_\Delta$. $\square$

We do not prove Theorem 2.2 as stated. Instead, it will be more convenient to prove a stronger result which we shall state (and prove) in Section 3.

Stacey and Weidl [104] also asked what happens in the context of colourings using finitely many colours. More precisely, they raised the following question: do there exist natural numbers $m \in \mathbb{N}$ with the property that for all sufficiently large $k \in \mathbb{N}$, $m \in \mathcal{G}_\Delta$ for every k-colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$? Observe that any such natural number m, assuming one exists, must be of the form $\binom{n}{2}$ or $\binom{n}{2} + 1$ for some natural number $n \geq 2$. One can see this by considering the family of 'small-rainbow colourings' of the complete graph on $\mathbb{N}$ which colour all the edges of some finite complete subgraph with distinct colours and all the remaining edges with a single colour not used in the finite (rainbow coloured) complete subgraph. On the other hand, when m is of the form $\binom{n}{2}$ or $\binom{n}{2} + 1$ for some natural number $n \geq 2$, we have the following positive result.

THEOREM 2.4. *For all* $n \in \mathbb{N}$, *there exists a natural number* $C = C(n)$ *such that for any* k*-colouring* $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ *with* $k \geq C$, *both* $\binom{n}{2}, \binom{n}{2} + 1 \in \mathcal{G}_\Delta$.

It turns out that the techniques used to prove Theorem 2.2 also allow us to prove a finitary version of the same theorem. In Section 4, we present this finitary result and use it prove Theorem 2.4 in a slightly stronger form. We conclude with some discussion in Section 5.

3. Proof of the main theorem

To prove Theorem 2.2, it will be more convenient to work with general infinite graphs. By an *infinite graph*, we mean a graph whose vertex set is $\mathbb{N}$ which has infinitely many edges.

It will be helpful to establish a few notational conveniences. Given an infinite graph G and an infinite-colouring $\Delta : G \twoheadrightarrow \mathbb{N}$ of the edges of G, for a subset X of $\mathbb{N}$, we shall write $\gamma_G(X)$, or just $\gamma(X)$ when both the colouring and the graph in question are clear from the context, for the number of distinct colours attained by Δ on $G[X]$, the subgraph of G induced by X; if H is a subgraph of G, we write $\gamma_H(\mathrm{X})$ for the the number of distinct colours attained by Δ on $H[X]$. For disjoint subsets X and Y, write $\gamma(X, Y)$ for the number of distinct colours in the induced bipartite subgraph between X and Y in G. Also, for a vertex $v \in \mathbb{N}$, we shall write $\gamma(v)$ for $\gamma(\{v\}, \mathbb{N} \setminus \{v\})$, the number of distinct colours of the edges incident to v in G.

We define the set $\mathcal{G}_\Delta$ for an infinite-colouring $\Delta : G \twoheadrightarrow \mathbb{N}$ of an infinite graph G in the obvious way by setting

$$\mathcal{G}_\Delta = \{\gamma_G(X) : X \subset \mathbb{N}\}.$$

In a general graph G, we say that X is *rainbow coloured in G* if $G[X]$ is a complete subgraph of G which is rainbow coloured. We say that X is *star coloured (with centre v) in G* if there is a vertex $v \in X$ such that $G[X \setminus \{v\}]$ is either an independent set or a 1-coloured complete graph, and all the edges between v and $X \setminus \{v\}$ are present and have distinct colours, which are also all different from the colour of $G[X \setminus \{v\}]$ in the case where $X \setminus \{v\}$ does not induce an independent set. The following result easily implies Theorem 2.2.

THEOREM 3.1. *For every infinite-colouring $\Delta : G \twoheadrightarrow \mathbb{N}$ of an infinite graph G, either*

(1) $\mathcal{G}_\Delta = \mathbb{N}$, or

(2) there exists an infinite rainbow coloured subset of $\mathbb{N}$; or

(3) there exists an infinite star coloured subset of $\mathbb{N}$.

For any finite set of colours $\mathcal{S}$, note that if we delete all the edges of an infinite graph G which are coloured with a colour from $\mathcal{S}$ by an infinite-colouring Δ of the edges of G, the resulting graph H is infinite and the restriction of Δ to H is an infinite-colouring. This makes the statement of Theorem 3.1 more amenable to induction than that of Theorem 2.2 and motivates the stronger statement of Theorem 3.1.

Fix an infinite-colouring $\Delta : G \twoheadrightarrow \mathbb{N}$ of an infinite graph G and note that if we have a partition $X = X_1 \cup X_2 \cup \ldots X_n$ of a subset X of $\mathbb{N}$, then

$$\sum_{1 \le i \le n} \gamma(X_i) + \sum_{1 \le i < j \le n} \gamma(X_i, X_j) \ge \gamma(X).$$

Consequently, if $\gamma(X) = \infty$, then at least one of the terms on the left is infinite; we shall make use of this fact repeatedly.

Next, we state a technical lemma about 'almost bipartite colourings' which will be useful in proving Theorem 3.1.

LEMMA 3.2. *Let G be an infinite graph and suppose that an infinite-colouring $\Delta : G \twoheadrightarrow \mathbb{N}$ of G is such that*

(1) $\gamma(v) < \infty$ for all $v \in \mathbb{N}$, and

(2) there is a partition of $\mathbb{N} = A \cup B$ such that $\gamma(A) < \infty$, $\gamma(B) < \infty$ and $\gamma(A, B) = \infty$.

Then for every natural number m, there exists a subset X of $\mathbb{N}$ such that $X \cap A \neq \varnothing$, $X \cap B \neq \varnothing$ and $\gamma(X) = m$.

Our strategy for proving both Theorem 3.1 and Lemma 3.2 is to inductively construct a set X for which $\gamma_G(X) = m$. To do this, we shall first delete some edges from G to get a new infinite graph H so that the restriction of Δ to H is also an infinite-colouring. We then inductively find a set Y with $\gamma_H(Y) = l$ for a suitably chosen $l < m$. Finally, we use the deleted edges in conjunction with Y to obtain X.

We first prove Lemma 3.2 and then show how to deduce Theorem 3.1 from it.

Proof of Lemma 3.2. Before we begin, let us note some consequences of our assumptions about the colouring Δ. Since $\gamma(v) < \infty$ for all $v \in \mathbb{N}$ and $\gamma(A, B) = \infty$, both A and B must be infinite. Furthermore, observe that if $\gamma(U) = \infty$ for some $U \subset \mathbb{N}$, then since $\gamma(A) < \infty$ and $\gamma(B) < \infty$, both $U \cap A$ and $U \cap B$ must be infinite.

We proceed by induction on m. The result is trivial for $m = 1$. Assuming the result for all $l < m$, we shall prove the result for m.

Pick an edge uv such that $u \in A$ and $v \in B$ and say that the colour of the edge is c. We know that $\gamma(u) < \infty$. We may assume, relabeling colours if necessary, that the colours of the edges incident to u are $1, \dots, \gamma(u)$. Consider the partition

$$\mathbb{N} \setminus \{u\} = U_0 \cup U_1 \cup \cdots \cup U_{\gamma(u)}$$

where U_0 is the set of vertices not adjacent to u in G and for $1 \leq i \leq \gamma(u)$, U_i is the set of all vertices that are joined to u by an edge of colour i. By considering the following three cases, we first show that we may assume that $\gamma(U_0) = \infty$.

Case 1: $\gamma(U_i) = \infty$ for some $i \neq 0$. We begin by observing (see Figure 2) that

$$\gamma(U_i \cap A) + \gamma(U_i \cap B) + \gamma(U_i \cap A, U_i \cap B) \geq \gamma(U_i).$$

Since $\gamma(U_i \cap A) \leq \gamma(A) < \infty$ and $\gamma(U_i \cap B) \leq \gamma(B) < \infty$, we conclude that $\gamma(U_i \cap A, U_i \cap B) = \infty$.

Let H be the infinite subgraph of $G[U_i]$ obtained by deleting all the edges of $G[U_i]$ of colour i. Then there exists, by the induction hypothesis, a subset Y of U_i such that $Y \cap (U_i \cap A) \neq \varnothing$, $Y \cap (U_i \cap B) \neq \varnothing$ and $\gamma_H(Y) = m - 1$.

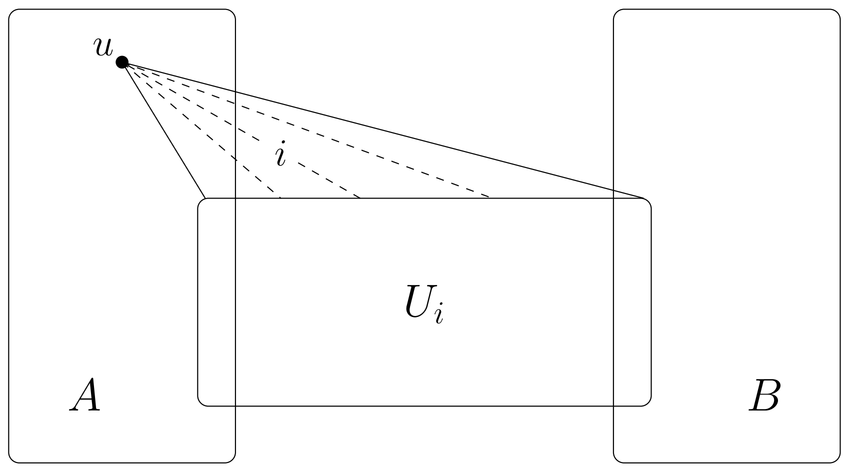

Figure 2. Case 1.

Observe that all the edges between u and $Y \subset U_i$ are coloured i in G. Since the colour i is not counted by γ_H, we see that $\gamma_G(Y \cup \{u\}) = m$. Therefore $X = Y \cup \{u\}$ is the required subset since $X \cap A \neq \varnothing$ and $X \cap B \neq \varnothing$.

Case 2: $\gamma(U_i, U_j) = \infty$ **for some** $0 < i < j$**.** Observe (see Figure 3) that $\gamma(U_i \cap A, U_j \cap A) \leq \gamma(A) < \infty$ and $\gamma(U_i \cap B, U_j \cap B) \leq \gamma(B) < \infty$. So we must either have $\gamma(U_i \cap A, U_j \cap B) = \infty$ or $\gamma(U_i \cap B, U_j \cap A) = \infty$. Without loss of generality, assume that $\gamma(U_i \cap A, U_j \cap B) = \infty$.

If $m \geq 3$, we may assume that the result holds for $m - 2$. Let H be the infinite subgraph of $G[(U_i \cap A) \cup (U_j \cap B)]$ obtained by deleting edges of colour i and j from $G[(U_i \cap A) \cup (U_j \cap B)]$. Then there exists, by the induction hypothesis, a subset Y of $(U_i \cap A) \cup (U_j \cap B)$ such that $Y \cap (U_i \cap A) \neq \varnothing$, $Y \cap (U_j \cap B) \neq \varnothing$ and $\gamma_H(Y) = m - 2$. Since $Y \subset U_i \cup U_j$, all the edges between u and Y in G are coloured either i or j and as $Y \cap U_i \neq \varnothing$ and $Y \cap U_j \neq \varnothing$, edges of both colours are present. Since both colours i and j are not counted by γ_H, it follows that $\gamma_G(Y \cup \{u\}) = m$. Clearly, $Y \cap A \neq \varnothing$ and $Y \cap B \neq \varnothing$ and therefore $X = Y \cup \{u\}$ is the required subset.

Now suppose that $m = 2$. Since $\gamma(w) < \infty$ for all $w \in \mathbb{N}$, we can greedily find an infinite matching $M = \{a_1 b_1, a_2 b_2, \dots\}$ between $U_i \cap A$ and $U_j \cap B$ in G such that each edge of the matching has a distinct colour; indeed, when we

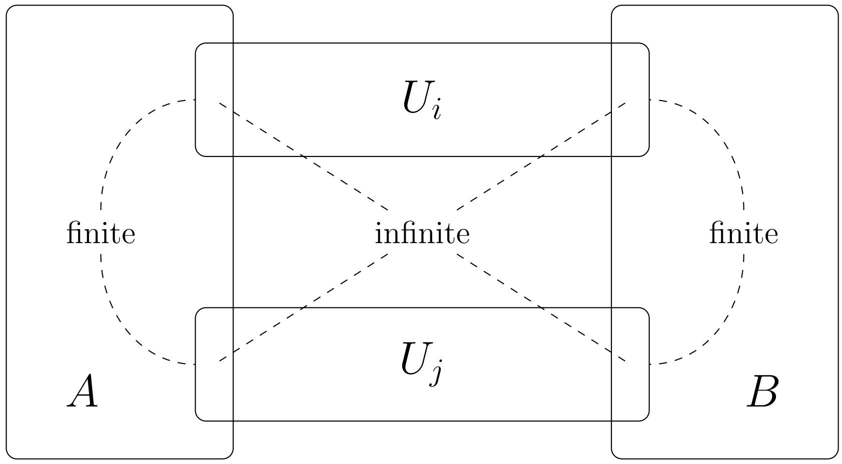

FIGURE 3. Case 2.

choose an edge and delete the edges incident to the ends of the chosen edge, we only loose finitely many colours from our graph. If a_k and b_l are not adjacent in G for some $k, l \in \mathbb{N}$, then $X = \{u, a_k, b_l\}$ is 2-coloured. So we may suppose that for each $k, l \in \mathbb{N}$, a_k is adjacent to b_l in G.

Since $\gamma(\{a_1, a_2, \dots\}) < \infty$ it follows from Ramsey's Theorem that there exists a subset $\{a'_1, a'_2, \dots\}$ of $\{a_1, a_2, \dots\}$ which either induces an independent set or a 1-coloured complete graph. Let a'_k be matched to the vertex b'_k in M and let c_k denote the colour of the edge $a'_k b'_k$.

If $\{a'_1, a'_2, \dots\}$ is an independent set in G, then since $\gamma(a'_1) < \infty$, there exist $s, t \in \mathbb{N}$ such that $a'_1 b'_s$ and $a'_1 b'_t$ have the same colour, say d. By our choice of M, $c_s \neq c_t$. Hence, at least one of c_s or c_t, say c_s, is not equal to d. Then it is easy to check that $X = \{a'_1, a'_s, b'_s\}$ is the required subset.

If $\{a'_1, a'_2, \dots\}$ induces a complete graph of colour d in G, we may assume (by discarding the edge $a'_1 b'_1$ and relabelling the remaining vertices if necessary) that c_1, the colour of the edge $a'_1 b'_1$, is not equal to d. Since $\gamma(b'_1) < \infty$, there exist $s, t \in \mathbb{N}$ such that $\Delta(a'_s b'_1) = \Delta(a'_t b'_1)$. If $\Delta(a'_s b'_1) = d$, then we may take $X = \{a'_1, a'_s, b'_1\}$. On the other hand, if $\Delta(a'_s b'_1) \neq d$, then $X = \{a'_s, a'_t, b'_1\}$ is the required subset.

Case 3: $\gamma(U_0, U_i) = \infty$ for some $i \neq 0$. We argue as we did in Case 2. We may assume that $\gamma(U_0 \cap A, U_i \cap B) = \infty$. Let H be the infinite subgraph of $G[(U_0 \cap A) \cup (U_i \cap B)]$ obtained by deleting all the edges of colour i from $G[(U_0 \cap A) \cup (U_i \cap B)]$.

By the induction hypothesis, there exists a subset Y of $(U_0 \cap A) \cup (U_i \cap B)$ such that $Y \cap (U_0 \cap A) \neq \varnothing$, $Y \cap (U_i \cap B) \neq \varnothing$ and $\gamma_H(Y) = m - 1$. As before, every edge between u and Y is coloured i in G (and u is adjacent to at least one vertex of Y since $Y \cap (U_i \cap B) \neq \varnothing$). Since the colour i is not counted by γ_H, it follows that $\gamma_H(Y \cup \{u\}) = m$. Hence, $X = Y \cup \{u\}$ is the required subset.

Hence, we may now assume that $\gamma(U_0) = \infty$. Since $\gamma(U_0) = \infty$, U_0 clearly meets both A and B in infinitely many vertices. We consider the graph induced by $U_0 \cup \{v\}$ and let V_0 be the set of those vertices of U_0 not adjacent to v in $G[U_0 \cup \{v\}]$. Since $\gamma(v) < \infty$, we have a partition of $U_0 \setminus V_0 = V_1 \cup \cdots \cup V_n$, with $n \leq \gamma(v)$, based on the colour of the edge joining a given vertex of $U_0 \setminus V_0$ to the vertex v. Applying the same argument as in Cases 1, 2 and 3 (which depended only on the vertex u and not on v) to the vertex v in $G[U_0 \cup \{v\}]$, we see that we are done unless $\gamma(V_0) = \infty$.

In this case, we consider the partition $V_0 = (V_0 \cap A) \cup (V_0 \cap B)$. Note that $\gamma(V_0 \cap A) < \infty$, $\gamma(V_0 \cap B) < \infty$ and $\gamma(V_0 \cap A, V_0 \cap B) = \infty$. Recall that we chose $u \in A$ and $v \in B$ such that the edge uv has colour c. Let H be the infinite subgraph of $G[V_0]$ obtained by deleting edges of colour c from $G[V_0]$.

By the induction hypothesis, there is a subset Y of V_0 such that $\gamma_H(Y) = m - 1$. Observe that uv has colour c and furthermore, u and v are not adjacent to any of the vertices of Y. Since the colour c is not counted by γ_H, we see that $\gamma_G(Y \cup \{u, v\}) = m$. Therefore $X = Y \cup \{u, v\}$ is the required subset since clearly, $X \cap A \neq \varnothing$ and $X \cap B \neq \varnothing$. This completes the proof. $\square$

We are now in a position to deduce Theorem 3.1 from Lemma 3.2.

Proof of Theorem 3.1. Let $\Delta : G \twoheadrightarrow \mathbb{N}$ be an infinite-colouring of an infinite graph G. We shall prove by induction on m that if G contains no infinite rainbow coloured or star coloured subset, then $m \in \mathcal{G}_\Delta$ for each $m \in \mathbb{N}$. The result is trivial for $m = 1$. Now suppose that $m \geq 2$. We shall inductively find a subset X of $\mathbb{N}$ with $\gamma(X) = m$.

If $\gamma(v) = \infty$ for some vertex $v \in \mathbb{N}$, then we can find an infinite subset $U = \{u_1, u_2, \dots\}$ of $\mathbb{N}$ such that the edges vu_i and vu_j have distinct colours for all $i \neq j$. Applying Theorem 2.1 to the restriction of Δ to $G[U]$ by colouring non-edges with a new colour, we can find an infinite subset $W = \{w_1, w_2, \dots\}$ of U such that W is either an independent set, 1-coloured, rainbow coloured or lexically coloured. We are done if W were rainbow coloured. If W is either an independent set or 1-coloured, it is clear that $W \cup \{v\}$ is star coloured with centre v. If W is lexically coloured, then it is easy to check that $\mathcal{G}_\Delta = \mathbb{N}$; indeed to find an m-coloured subgraph, we consider the subgraph induced by the first $m+1$ vertices of W and note that this subgraph induces exactly m colours unless one of these m colours corresponds to the new colour corresponding to non-edges, in which case we may take the subgraph induced by the first $m+2$ vertices of W.

So we may assume that $\gamma(v) < \infty$ for all $v \in \mathbb{N}$. Pick an edge uv of G, and say that the colour of the edge is c. We may suppose that the colours of the edges incident to u are $1, \dots, \gamma(u)$. Consider the partition $\mathbb{N} \backslash \{u\} = U_0 \cup U_1 \cup \dots \cup U_{\gamma(u)}$ where U_0 is the set of vertices not adjacent to u in G and for $1 \leq i \leq \gamma(u)$, U_i is the set of all vertices that are joined to u by an edge of colour i. Since $\gamma(\mathbb{N}) = \infty$, by the pigeonhole principle, we must either have $\gamma(U_i) = \infty$ for some i, or $\gamma(U_i, U_j) = \infty$ for some $i \neq j$. We distinguish the following cases.

Case 1: $\gamma(U_i) < \infty$ for all $0 \leq i \leq \gamma(u)$. Since $\gamma(\mathbb{N}) = \infty$, it must be the case that $\gamma(U_i, U_j) = \infty$ for some $i \neq j$. Applying Lemma 3.2 to the restriction of Δ to $G[U_i \cup U_j]$, we find a subset X of $U_i \cup U_j$ such that $\gamma_G(X) = m$.

Case 2: $\gamma(U_i) = \infty$ for some $i \neq 0$. Let H be the infinite subgraph of $G[U_i]$ obtained by deleting all the edges of colour i from $G[U_i]$. Clearly,

$\gamma_H(w) < \infty$ for all $w \in U_i$. So H contains no infinite subset which is rainbow or star coloured. By the induction hypothesis, there is a subset Y of U_i such that $\gamma_H(Y) = m - 1$. Observe that all the edges between u and $Y \subset U_i$ have colour i, and since the colour i is not counted by γ_H, we see that $\gamma_G(Y \cup \{u\}) = m$. Therefore $X = Y \cup \{u\}$ is the required subset.

Case 3: $\gamma(U_0) = \infty$. Let V_0 be the set of those vertices of U_0 not adjacent to v in G. Since $\gamma(v) < \infty$, we have a partition of $U_0 \setminus V_0 = V_1 \cup \cdots \cup V_n$, with $n \leq \gamma(v)$, based on the colour of the edge joining a given vertex of $U_0 \setminus V_0$ to the vertex v. Applying the same argument as in Cases 1 and 2 to the vertex v, we see that we are done unless $\gamma(V_0) = \infty$. In this case, we consider the infinite subgraph H of $G[V_0]$ obtained by deleting all the edges of colour c from $G[V_0]$.

The fact that $\gamma_G(w) < \infty$ for all $w \in \mathbb{N}$ implies that $\gamma_H(w) < \infty$ for all $w \in V_0$. So H has no infinite rainbow or star coloured subset. By the induction hypothesis, there is a subset Y of V_0 such that $\gamma_H(Y) = m - 1$. Observe that uv has colour c and there are no edges between $\{u, v\}$ and $Y \subset V_0 \subset U_0$ in G. Since the colour c is not counted by γ_H, it follows that $\gamma_G(Y \cup \{u, v\}) = m$. Therefore $X = Y \cup \{u, v\}$ is the required subset. This completes the proof. $\square$

4. Extensions and applications

In this section, we shall first describe a finitary analogue of Theorem 2.2. We then use this to prove Theorem 2.4. For us, a *countable set* is a set that is either finite or countably infinite.

4.1. Finitary extensions. We can prove a version of Theorem 2.2 for colourings (of finite or infinite complete graphs) that use only finitely many colours.

THEOREM 4.1. *For all $n \in \mathbb{N}$, there exists a natural number $K = K(n)$ such that for every k-colouring $\Delta : V^{(2)} \twoheadrightarrow [k]$ of the complete graph on a countable set V with $k \geq K$ colours, either*

(1) there is an m-coloured complete subgraph for every $m \in [n]$, or

(2) there exists a rainbow coloured complete subgraph on n vertices, or

(3) there exists a star coloured complete subgraph on n vertices. $\square$

This result can be proved by arguments similar to those used to prove Theorem 2.2. There are two essential differences. First, as opposed to Theorem 2.1, we use the following extension of the theorem proved by Erdős and Rado, to colourings of *finite* complete graphs with an arbitrary set of colours.

THEOREM 4.2. *For every $n \in \mathbb{N}$, and every colouring Δ of the complete graph on a sufficiently large countable set V, there exists a subset X of V of size at least n such that either*

(1) X is 1-coloured, or

(2) X is rainbow coloured, or

(3) X is lexically coloured. $\square$

Second, in the place of Lemma 3.2, we use the following finitary analogue which is proved in the same way as the lemma.

LEMMA 4.3. *For all $m, d \in \mathbb{N}$, there exists a natural number $L = L(m, d)$ with the following property: for every colouring Δ of a graph G on a countable set V such that*

(1) $\gamma(v) < d$ for all $v \in V$, and

(2) there is a partition of $V = A \cup B$ such that $\gamma(A) < d$, $\gamma(B) < d$ and $\gamma(A, B) \geq L$,

there exists a subset X of V such that $X \cap A \neq \varnothing$, $X \cap B \neq \varnothing$ and $\gamma(X) = m$. $\square$

4.2. Applications. Theorem 2.4 may be deduced from Theorem 4.1. Recall that Theorem 2.4 says for any natural number $n \in \mathbb{N}$, both $\binom{n}{2}, \binom{n}{2}+1 \in \mathcal{G}_\Delta$ for any colouring Δ of the complete graph on $\mathbb{N}$ using a finite, but sufficiently large number of colours.

We prove two propositions which, taken together, imply the result. The first is an easy corollary of Theorem 4.1

PROPOSITION 4.4. *For all $n \in \mathbb{N}$, there exists a natural number $C_1 = C_1(n)$ such that for any k-colouring $\Delta : V^{(2)} \twoheadrightarrow [k]$ of the complete graph on a countable set V with $k \geq C_1$ colours, $\binom{n}{2} \in \mathcal{G}_\Delta$.*

PROOF. Take $C_1(n) = K(\binom{n}{2})$, where K is as guaranteed by Theorem 4.1.

□

The next proposition is perhaps not as straightforward.

PROPOSITION 4.5. *For all $n \in \mathbb{N}$, there exists a natural number $C_2 = C_2(n)$ with the property that for all $k \geq C_2$, there exists a natural number $D_{k,n}$ such that for any k-colouring $\Delta : V^{(2)} \twoheadrightarrow [k]$ of the complete graph on a countable set V with $k \geq C_2$ colours, $\binom{n}{2} + 1 \in \mathcal{G}_\Delta$, provided $|V| \geq D_{k,n}$.*

PROOF. For $n = 2$, it is an easy exercise to check that the result is true with $C_2(2) = 2$ and $D_{k,2} = R(k+1; k)$ where $R(k+1; k)$ is the Ramsey number for finding a 1-coloured copy of a complete graph on $k + 1$ vertices when using k colours. Indeed, given a k-colouring of a complete graph on V with $|V| \geq R(k+1; k)$, we first find a maximal 1-coloured set $X \subset V$ of size at least $k+1$ coloured say, blue. Consider a vertex $v \notin X$. If v is joined to some vertex u of X by a blue edge, we may extend $\{v, u\}$ to a 2-coloured triangle using the maximality of X. If no edge between v and X is coloured blue, then since $|X| \geq k + 1$, there are two vertices in X which are joined to v by edges of the same colour, again allowing us to find a 2-coloured triangle.

For $n \geq 3$, let $s = n^4$. We claim that $C_2(n) = K(s)$ will do, where K is the constant guaranteed by Theorem 4.1. For $k \geq C_2(n)$, we take $D_{k,n} = k^s + s + 1$. Now, suppose that $\Delta : V^{(2)} \twoheadrightarrow [k]$ is a k-colouring and $|V| \geq D_{k,n}$. Then, by our choice of $C_2(n)$, either

(1) there is an m-coloured complete subgraph for every $m \in [s]$, or

(2) there exists a rainbow coloured complete subgraph on s vertices, or

(3) there exists a star coloured complete subgraph on s vertices.

Note that a star coloured complete subgraph on s vertices contains an m-coloured complete subgraph for $2 < m \leq s$. Since $2 < \binom{n}{2} + 1 \leq s$, we are done unless there exists a rainbow coloured complete subgraph on s vertices. Hence, suppose that the complete subgraph on the vertex set $S = \{u_1, u_2, \ldots, u_s\}$ is rainbow coloured. For each $x \in V \setminus S$, there are k^s possible values for the s-tuple $(\Delta(xu_1), \Delta(xu_2), \ldots, \Delta(xu_s))$. Since, $|V \setminus S| \geq D_{k,n} - s > k^s$, we can find vertices $x, y \in V \setminus S$ such that

$$(\Delta(xu_1), \Delta(xu_2), \ldots, \Delta(xu_s)) = (\Delta(yu_1), \Delta(yu_2), \ldots, \Delta(yu_s)).$$

We claim that there is a subset $T \subset S$ of size $t = n^2$ such that for all $u \in T$, $\Delta(xu) \notin \Delta(T^{(2)})$. Consider the set

$$A = \{(u, T) : u \in T \subset S,\ |T| = t,\ \Delta(xu) \in \Delta(T^{(2)})\}.$$

As S is rainbow coloured, there is at most one edge ab in $S^{(2)}$ of colour $\Delta(xu)$ for each $u \in S$. If (u, T) is in A, then we must have $a, b \in T$. So for each $u \in S$, there are at most $\binom{s-2}{t-2}$ sets T such that $(u, T) \in A$. Therefore, $|A| \leq s\binom{s-2}{t-2} < \binom{s}{t}$ since $s = t^2$. Hence there exists a T which does not appear in A and the claim follows.

Hence, there is indeed a subset T of S of size $t = n^2$ such that $\Delta(xu) \notin \Delta(T^{(2)})$ for all $u \in T$. Let $\mathcal{Q} = \{\Delta(xu) : u \in T\}$. If $|\mathcal{Q}| < n$, then as $|T| = n^2$, there are vertices $v_1, v_2, \ldots, v_n$ in T such that

$$\Delta(xv_1) = \Delta(xv_2) = \cdots = \Delta(xv_n).$$

Since this colour $\Delta(xv_1)$ is not an element of $\Delta(T^{(2)})$, we conclude that the set $\{x, v_1, v_2, \ldots, v_n\}$ is $(\binom{n}{2} + 1)$-coloured.

So we may assume that $|\mathcal{Q}| \geq n$. Then there is a subset $U \subset T$ of size n such that the colours $\Delta(xu)$ are distinct for all $u \in U$. Since $U \subset T$, the colour

$\Delta(xu)$ is not an element of $\Delta(U^{(2)})$ for each $u \in U$. We hence conclude that $U \cup \{x\}$ is rainbow coloured.

Recall that there is a vertex $y \neq x$ in $V \setminus S$ such that $\Delta(xu) = \Delta(yu)$ for all $u \in S$. Since at most one edge e in $(U \cup \{x\})^{(2)}$ is coloured with the same colour as the edge xy, by removing the endpoint of e which lies in U if necessary, we can find a subset U' of U of size $n-1$ such that $\Delta(xy)$ is not an element of $\Delta((U' \cup \{x\})^{(2)})$. Then $U' \cup \{x, y\}$ is $(\binom{n}{2}+1)$-coloured since $U' \cup \{x\}$ and $U' \cup \{y\}$ are rainbow coloured sets of size n using the same set of colours. $\square$

It is easy to see that, taken together, Corollary 4.4 and Theorem 4.5 imply Theorem 2.4.

The following corollary of Lemma 4.3 about finding m-coloured complete bipartite subgraphs might be of independent interest.

Corollary 4.6. *For all $m \in \mathbb{N}$, there exists a natural number $B = B(m)$ such that if $\Delta : U \times V \twoheadrightarrow [k]$ is a k-colouring of the complete bipartite graph between two countable sets U and V with $k \geq B$ colours, then there exist $X \subset U$ and $Y \subset V$ such that the complete bipartite subgraph between by X and Y is m-coloured.*

Proof. It is easy to check that it suffices to take $B(m) = L(m, m)$, where L is the constant guaranteed by Lemma 4.3. $\square$

5. Concluding remarks

We conclude by mentioning two questions that would merit further study. First, the problem of determining for each $k \in \mathbb{N}$, which natural numbers $m \in \mathbb{N}$ are guaranteed to belong to $\mathcal{G}_\Delta$ for every k-colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ is quite interesting; while we have taken a few steps towards this in this chapter, the full question is still far from being resolved. Second, it would be reasonable to ask the questions considered here for uniform hypergraphs. However, even in the case of $\mathbb{N}^{(3)}$, it is not immediately clear to us what the canonical structures analogous to the rainbow coloured and star coloured complete graphs should be.

CHAPTER 3

Infinite exactly m-coloured complete graphs

1. Introduction

In the last chapter, we investigated, for a given colouring of the complete graph on the natural numbers, the properties of the set of those natural numbers m for which there exists an m-coloured complete subgraph. In this chapter, we shall study how the situation changes when one is interested in finding a 'large' m-coloured complete subgraph.

We begin by briefly recalling some of the definitions from the previous chapter. Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$ be a surjective colouring of the edges of the complete graph on $\mathbb{N}$ with an arbitrary set of colours $\mathcal{C}$. If the set of colours $\mathcal{C}$ is infinite, we say that Δ is an *infinite-colouring* and if $\mathcal{C}$ is finite, we say that Δ is a k*-colouring* if $|\mathcal{C}| = k$. Given a colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$ of the complete graph on $\mathbb{N}$, we say that a subset X of $\mathbb{N}$ is (*exactly*) m*-coloured* if $\Delta(X^{(2)})$, the set of values attained by Δ on the edges with both endpoints in X, has size exactly m. Let $\gamma_\Delta(X)$, or $\gamma(X)$ in short, denote the size of the set $\Delta(X^{(2)})$; in other words, every set X is $\gamma(X)$-coloured.

In the last chapter, we studied, for a colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$ of the complete graph on $\mathbb{N}$, the properties of the set

$$\mathcal{G}_\Delta = \{\gamma_\Delta(X) : X \subset \mathbb{N}\}.$$

In the context of Ramsey theory, one is usually interested in finding 'large' homogeneous structures with certain properties. With this in mind, for a colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$, we define

$$\mathcal{F}_\Delta = \{\gamma_\Delta(X) : X \subset \mathbb{N} \text{ such that } X \text{ is } \textit{infinite}\}.$$

When Δ is an infinite-colouring, it might so happen that for each infinite subset X of $\mathbb{N}$, the set $\Delta(X^{(2)})$ is infinite; consequently, it is only really meaningful to study the set $\mathcal{F}_\Delta$ in the case of colourings using finitely many colours. This question of finding infinite m-coloured complete subgraphs was first considered by Erickson [52]. If $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ is a k-colouring of the edges of the complete graph on the natural numbers, then clearly $k \in \mathcal{F}_\Delta$ as Δ is surjective, and Ramsey's Theorem tells us that $1 \in \mathcal{F}_\Delta$. Erickson [52] noted that a fairly straightforward application of Ramsey's Theorem enables one to show that $2 \in \mathcal{F}_\Delta$. Erickson conjectured however that with the exception of 1, 2 and k, no other elements are guaranteed to be in $\mathcal{F}_\Delta$.

Conjecture 1.1. *Let $k, m \in \mathbb{N}$ with $k > m > 2$. Then there exists a k-colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ such that $m \notin \mathcal{F}_\Delta$.*

Stacey and Weidl [104] settled this conjecture in the case where k is much bigger than m. More precisely, for any $m > 2$, they showed that there exists a constant C_m such that if $k > C_m$, then there is a k-colouring Δ such that $m \notin \mathcal{F}_\Delta$.

Erickson's conjecture, if true, would suggest that it is hopeless to look for particular values in the set $\mathcal{F}_\Delta$ given a k-colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$. It is natural then to consider other properties of the set $\mathcal{F}_\Delta$. The first question which arises is that of the set of possible sizes of $\mathcal{F}_\Delta$. Since $\mathcal{F}_\Delta \subset [k]$, it follows that $|\mathcal{F}_\Delta| \leq k$ and it is easy to see that equality is in fact possible. Things are not so clear when we turn to the question of lower bounds. Let us define

$$\psi(k) = \min_{\Delta:\mathbb{N}^{(2)} \twoheadrightarrow [k]} |\mathcal{F}_\Delta|.$$

We are able to prove the following lower bound for $\psi(k)$.

Theorem 1.2. *Let $n \geq 2$ be the largest natural number such that $k \geq \binom{n}{2}+1$. Then $\psi(k) \geq n$.*

It is not hard to check that Theorem 1.2 is tight when $k = \binom{n}{2} + 1$ for some $n \geq 2$. To this end, we consider the 'small-rainbow colouring' Δ which colours all the edges with both endpoints in $[n]$ with $\binom{n}{2}$ distinct colours and all the remaining edges with the one colour that has not been used so far. Clearly $\mathcal{F}_\Delta = \{\binom{i}{2} + 1 : i \leq n\}$, and so Theorem 1.2 is best possible for infinitely many values of k.

Turning to the question of upper bounds for ψ, the small-rainbow colouring demonstrates that $\psi(k) = O(\sqrt{k})$ for infinitely many values of k. When k is not of the form $\binom{n}{2} + 1$, there are two obvious ways of generalising the small-rainbow colouring described above: we could replace the rainbow coloured clique in the construction either with a disjoint union of cliques, or with a clique along with an extra vertex attached to some vertices of the clique. It is not hard to check that both these generalisations fail to give us good upper bounds for $\psi(k)$ for general k; in particular, we are unable to decide if $\psi(k) = o(k)$ for all $k \in \mathbb{N}$. However, by considering colourings that colour all the edges of a small complete bipartite graph with distinct colours (as opposed to a small complete graph) and making use of some number theoretic estimates of Tenenbaum [105] and Ford [56], we get reasonably close to such a statement.

Theorem 1.3. *There exists a subset A of the natural numbers of asymptotic density one such that for all $k \in A$,*

$$\psi(k) = O\left(\frac{k}{(\log\log k)^{\delta}(\log\log\log k)^{3/2}}\right)$$

where $\delta = 1 - \frac{1+\log\log 2}{\log 2} \approx 0.086 > 0$.

The rest of this chapter is organised as follows. In the next section, we prove Theorem 1.2. We remark that we do not prove Theorem 1.2 as stated. Instead, we prove a stronger structural result that implies the theorem. We postpone the statement of this result since it depends on a certain notion of homogeneity that we shall introduce in the next section. In Section 3, we describe how Theorem 1.3 follows from certain divisor estimates. We conclude by mentioning some open problems in Section 4.

2. Lower bounds

In this section, we prove Theorem 1.2 by proving a stronger structural result, namely Theorem 2.3.

We first introduce a notational convenience. Given a colouring Δ of $\mathbb{N}^{(2)}$, a vertex $v \in \mathbb{N}$, and a subset $X \subset \mathbb{N} \setminus \{v\}$, we say that a colour c is a *new colour from v into X* if some edge from v to X is coloured c by Δ and also, no edge of $X^{(2)}$ is coloured c by Δ. We write $N_\Delta(v, X)$, or just $N(v, X)$ when the colouring Δ in question is clear, for the set of new colours from v into X.

2.1. Proof of Theorem 1.2. Before we prove Theorem 1.2, we note that Erickson's argument showing that $2 \in \mathcal{F}_\Delta$ can be generalised to give a quick proof that $\psi(k) = \Omega(\log k)$.

Lemma 2.1. *Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ be a k-colouring and suppose $l \in \mathcal{F}_\Delta$ and $l < k$. Then there is an $m \in \mathcal{F}_\Delta$ such that $l + 1 \leq m \leq 2l$.*

Proof of Lemma 2.1. The collection of infinite l-coloured sets is non-empty and so by Zorn's lemma, there is a maximal infinite l-coloured set; let $X \subset \mathbb{N}$ be such a set. As $l < k$, $X \neq \mathbb{N}$. Pick $v \in \mathbb{N} \setminus X$. Note that $N(v, X) \neq \varnothing$ since otherwise $X \cup \{v\}$ is l-coloured, which contradicts the maximality of X.

If $|N(v, X)| \leq l$, then $X \cup \{v\}$ is m-coloured for some $l + 1 \leq m \leq 2l$. So suppose $|N(v, X)| \geq l + 1$. By the pigeonhole principle, there is an infinite subset Y of X such that all the vertices of Y are connected to v by edges of a single colour, say c.

We consider two cases. If $c \in N(v, X)$, we pick $l - 1$ vertices from X which are joined to v by edges coloured with $l - 1$ distinct colours from $N(v, X) \setminus \{c\}$. If on the other hand $c \notin N(v, X)$, we pick l vertices from X which are joined to v by edges coloured with l distinct colours from $N(v, X)$. Call this set of $l - 1$ or l vertices Z.

In both cases, it is easy to check that $Y \cup Z \cup \{v\}$ is m-coloured with $l+1 \leq m \leq 2l$ since the number of colours appearing in $Y \cup Z$ is between 1 and l and v introduces precisely l new colours into this set. $\square$

Note that Lemma 2.1, coupled with the fact that we always have $1 \in \mathcal{F}_\Delta$, implies that $\psi(k) \geq 1 + \log_2 k$. By applying Lemma 2.1 to the largest element of $\mathcal{F}_\Delta$ in $[2^n]$, we have the following corollary.

Corollary 2.2. *If $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ is a k-colouring and n is a natural number such that $k \geq 2^n + 1$, then $\mathcal{F}_\Delta \cap ([2^{n+1}] \setminus [2^n]) \neq \varnothing$.* $\square$

We shall show that for any k-colouring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ with $k \geq \binom{n}{2} + 1$ for some n, we can find n nested subsets $A_1 \subsetneq A_2 \subsetneq \cdots \subsetneq A_n$ of $\mathbb{N}$ such that $\Delta(A_1^{(2)}) \subsetneq \Delta(A_2^{(2)}) \subsetneq \cdots \subsetneq \Delta(A_n^{(2)})$. To do this, we introduce the notion of n-homogeneity on which our first structural result, Theorem 2.3, hinges.

For an ordered n-tuple $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, write $\widehat{X}_i$ for the set $X_1 \cup X_2 \cdots \cup X_i$. Given a colouring Δ, we call $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, with each X_i a non-empty subset of $\mathbb{N}$, *n-homogeneous with respect to Δ* if the following conditions are met:

(1) $X_i \cap X_j = \varnothing$ for $i \neq j$,
(2) X_1 is infinite and 1-coloured,
(3) $\Delta(\widehat{X}_1^{(2)}) \subsetneq \Delta(\widehat{X}_2^{(2)}) \subsetneq \cdots \subsetneq (\widehat{X}_n^{(2)})$,
(4) for each X_i with $2 \leq i \leq n$, every $v \in X_i$ satisfies

$$N(v, \widehat{X}_{i-1}) = \Delta\left(\widehat{X}_i^{(2)}\right) \setminus \Delta\left(\widehat{X}_{i-1}^{(2)}\right), \text{ and}$$

(5) $\gamma(\widehat{X}_n) \leq \binom{n}{2} + 1$.

Rather than proving Theorem 1.2, we prove the following stronger statement.

Theorem 2.3. *Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ be a k-colouring and suppose n is a natural number such that $k \geq \binom{n}{2} + 1$. Then there exists an n-homogeneous tuple with respect to Δ.*

Proof. We proceed by induction on n. The case $n = 1$ is Ramsey's Theorem. Suppose that $k \geq \binom{n+1}{2} + 1$ and assume inductively that at least one n-homogeneous tuple exists; let $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ be such a tuple. Note that $k \geq \binom{n+1}{2} + 1 > \binom{n}{2} + 1$. Since Δ is surjective and attains at most $\binom{n}{2} + 1$ different values inside $\widehat{X}_n$, clearly $\mathbb{N} \setminus \widehat{X}_n \neq \varnothing$. We consider two cases.

Case 1: $N(v, \widehat{X}_n) \neq \varnothing$ for some $v \in \mathbb{N} \setminus \widehat{X}_n$. If $|N(v, \widehat{X}_n)| \leq n$, then it is easy to check that $(X_1, X_2, \ldots, X_n, \{v\})$ is an $(n+1)$-homogeneous tuple and we are done. So, assume without loss of generality that $|N(v, \widehat{X}_n)| \geq n + 1$.

Let j be the smallest index such that $N(v, \widehat{X}_j) \neq \varnothing$. Since $N(v, \widehat{X}_n) \neq \varnothing$, this minimal index j exists. We now build our $(n+1)$-homogeneous tuple $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{n+1})$ as follows.

Set $Y_1 = X_1, Y_2 = X_2, \ldots, Y_{j-1} = X_{j-1}$. We define Y_j as follows. First, choose $c \in N(v, \widehat{X}_j)$; note that by the minimality of j, $N(v, \widehat{X}_{j-1}) = \varnothing$ and so all the edges between v and $\widehat{X}_j$ coloured c are actually edges between v and X_j. Take $Y_j \subset X_j$ to be the (non-empty) set of vertices $u \in X_j$ such that the edge between v and u is either coloured c or with a colour from $\Delta(\widehat{X}_j^{(2)})$ (and hence a colour not in $N(v, \widehat{X}_j)$). Note that if $j = 1$, we can always choose c such that Y_1 is an infinite subset of X_1.

Next, set $Y_{j+1} = \{v\}$. Now, note that the only colour from $\Delta(\widehat{Y}_{j+1}^{(2)})$ that might possibly occur in $N(v, \widehat{X}_n)$ is c. So we can now choose $v_1, v_2, \ldots, v_{n-j}$ from $X_n \cup X_{n-1} \cdots \cup X_{j+1} \cup (X_j \setminus Y_j)$ such that these $n - j$ vertices are joined to v by edges which are all coloured by distinct elements of $N(v, \widehat{X}_n) \setminus \{c\}$. Set $Y_{j+2} = \{v_1\}, Y_{j+3} = \{v_2\}, \ldots, Y_{n+1} = \{v_{n-j}\}$.

We claim that $\mathbf{Y}$ is an $(n+1)$-homogeneous tuple. Indeed, conditions (1) and (2) are obviously satisfied.

To check condition (3), first note that $\Delta(\widehat{Y}_1^{(2)}) \subsetneq \Delta(\widehat{Y}_2^{(2)}) \subsetneq \cdots \subsetneq \Delta(\widehat{Y}_{j-1}^{(2)})$ follows from the n-homogeneity of $\mathbf{X}$ since $Y_i = X_i$ for $1 \leq i \leq j - 1$. Also, $\Delta(\widehat{Y}_{j-1}^{(2)}) \subsetneq \Delta(\widehat{Y}_j^{(2)})$ since Y_j is a non-empty subset of X_j. Next, $\Delta(\widehat{Y}_j^{(2)}) \subsetneq \Delta(\widehat{Y}_{j+1}^{(2)})$ since v is joined to at least one vertex of Y_j by an edge coloured with

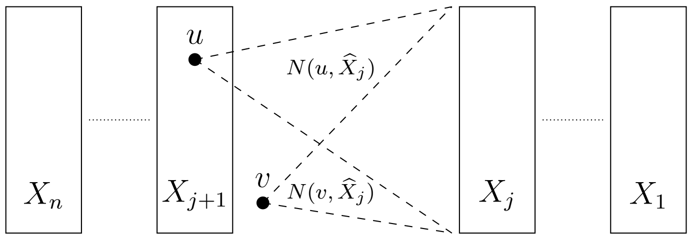

FIGURE 1. Inserting v into $\mathbf{X}$.

c and we know that c is a new colour from v into $\widehat{Y}_j$. Finally, $\Delta(\widehat{Y}_{j+1}^{(2)}) \subsetneq \Delta(\widehat{Y}_{j+2}^{(2)}) \subsetneq \cdots \subsetneq \Delta(\widehat{Y}_{n+1}^{(2)})$ because the vertices $v_1, v_2, \ldots, v_{n-j}$ are all joined to v by edges of distinct colours and none of these colours belong to $\Delta(\widehat{X}_n^{(2)})$. So condition (3) is also satisfied.

Condition (4) for each of $Y_1, Y_2, \ldots, Y_j$ is equivalent to the same condition for $X_1, X_2, \ldots, X_j$ respectively. Furthermore, condition (4) is also satisfied by each of $Y_{j+1}, Y_{j+2}, \ldots, Y_{n+1}$ since they each contain exactly one vertex.

Finally, we check condition (5). Clearly, $\Delta(\widehat{Y}_{n+1}^{(2)})$ is a subset of $\Delta(\widehat{X}_n^{(2)}) \cup T$ for some subset T of $N(v, \widehat{X}_n)$ of size at most n. Hence, we see that $\gamma(\widehat{Y}_{n+1}) \leq \binom{n}{2} + 1 + n = \binom{n+1}{2} + 1$.

Case 2: $N(v, \widehat{X}_n) = \varnothing$ **for every** $v \in \mathbb{N} \setminus \widehat{X}_n$**.** To deal with this case, we will need the following lemma.

LEMMA 2.4. *Let* $\mathbf{X}$ *be an* n*-homogeneous tuple and suppose* $N(v, \widehat{X}_n) = \varnothing$ *for some* $v \in \mathbb{N} \setminus \widehat{X}_n$*. Then, there either exists an* $(n+1)$*-homogeneous tuple* $\mathbf{Y}$*, or an* n*-homogeneous tuple* $\mathbf{Z}$ *such that* $Z_j = X_j \cup \{v\}$ *for some* $j \in [n]$*, and* $Z_i = X_i$ *for each* $1 \leq i \leq n$ *with* $i \neq j$*.*

PROOF. If $N(v, \widehat{X}_i) = \varnothing$ for $1 \leq i \leq n$, then $(X_1 \cup \{v\}, X_2, \ldots, X_n)$ is n-homogeneous and we have $\mathbf{Z}$ as required. Hence, let $j < n$ be the largest index such that $N(v, \widehat{X}_j) \neq \varnothing$. So by the definition of j, $N(v, \widehat{X}_i) = \varnothing$ for $j < i \leq n$. Let $\mathbf{Z} = (X_1, X_2, \ldots, X_j, X_{j+1} \cup \{v\}, X_{j+2}, \ldots, X_n)$ and let $\mathbf{Y} = (X_1, X_2, \ldots, X_j, \{v\}, X_{j+1}, X_{j+2}, \ldots, X_n)$; we claim that either $\mathbf{Z}$ is n-homogeneous or $\mathbf{Y}$ is $(n+1)$-homogeneous.

Consider a colour c that belongs to $N(v, \widehat{X}_j)$. Since $N(v, \widehat{X}_{j+1}) = \varnothing$, this means that c must occur in $\Delta(\widehat{X}^{(2)}_{j+1}) \setminus \Delta(\widehat{X}^{(2)}_j)$. But, by condition (4), for each $u \in X_{j+1}$, $N(u, \widehat{X}_j) = \Delta(\widehat{X}^{(2)}_{j+1}) \setminus \Delta(\widehat{X}^{(2)}_j)$. Hence, $N(v, \widehat{X}_j) \subset N(u, \widehat{X}_j)$ for $u \in X_{j+1}$.

We first show that if $N(v, \widehat{X}_j) = N(u, \widehat{X}_j)$ for $u \in X_{j+1}$, then $\mathbf{Z}$ is n-homogeneous. Conditions (1) and (2) are clearly satisfied by $\mathbf{Z}$. To see that conditions (3) and (4) hold, first note that these conditions hold for all the Z_i with $1 \leq i \leq j$ since $Z_i = X_i$ for $1 \leq i \leq j$. Both conditions hold for Z_{j+1} since we have assumed that $N(v, \widehat{X}_j) = N(u, \widehat{X}_j)$ for $u \in X_{j+1}$. Next, observe that since $N(v, \widehat{X}_i) = \varnothing$ for $j < i \leq n$, we have $N(u, \widehat{X}_{i-1}) = N(u, \widehat{X}_{i-1} \cup \{v\})$ for each $u \in X_i$ with $j + 1 < i \leq n$. From this observation, it follows that both conditions hold for all the Z_i with $j + 1 < i \leq n$. Finally, condition (5) holds since $N(v, \widehat{X}_n) = \varnothing$.

If we instead have $N(v, \widehat{X}_j) \subsetneq N(u, \widehat{X}_j)$ for $u \in X_{j+1}$, then we claim that $\mathbf{Y}$ is $(n + 1)$-homogeneous.

Clearly, conditions (1) and (2) are satisfied by $\mathbf{Y}$. To check conditions (3) and (4), we proceed as we did previously for $\mathbf{Z}$. First note that these conditions hold for all the Y_i such that $1 \leq i \leq j$ since $Y_i = X_i$ for $1 \leq i \leq j$. Both conditions hold for the Y_{j+1} since Y_{j+1} consists of a single vertex v and since $N(v, \widehat{X}_j) \neq \varnothing$. To see that condition (3) holds for Y_{j+2}, note that $\Delta(\widehat{Y}^{(2)}_{j+1}) \subsetneq \Delta(\widehat{Y}^{(2)}_{j+2})$ since $N(v, \widehat{X}_j) \subsetneq N(u, \widehat{X}_j)$ for $u \in X_{j+1}$. We know that $N(v, \widehat{X}_{j+1}) = \varnothing$. Hence, condition (4) also holds for Y_{j+2} since for any vertex $u \in Y_{j+2} = X_{j+1}$, we see that $N(u, \widehat{Y}_{j+1}) = N(u, \widehat{X}_j) \setminus N(v, \widehat{X}_j) = \Delta(\widehat{Y}^{(2)}_{j+2}) \setminus \Delta(\widehat{Y}^{(2)}_{j+1})$. Finally, both conditions also hold for all Y_i with $j + 2 < i \leq n + 1$. This follows from the fact that $N(u, \widehat{X}_{i-1} \cup \{v\}) = N(u, \widehat{X}_{i-1}) \neq \varnothing$ for each $u \in X_i$ with $j + 1 < i \leq n$. Finally, it is easy to see that condition (5) holds since $N(v, \widehat{X}_n) = \varnothing$. $\square$

We have assumed that $N(v, \widehat{X}_n) = \varnothing$ for each $v \in \mathbb{N} \setminus \widehat{X}_n$. Now, Δ is surjective, so there must exist two vertices v_1 and v_2 in $\mathbb{N} \setminus \widehat{X}_n$ such that the edge joining v_1 and v_2 is coloured with a colour c not in $\Delta(\widehat{X}_n^{(2)})$.

We apply Lemma 2.4 to $\mathbf{X}$ and v_1. If we find an $(n+1)$-homogeneous tuple $\mathbf{Y}$, we are done. So suppose that the lemma yields n-homogeneous tuple $\mathbf{Z}$. Then clearly $N(v_2, \widehat{Z}_n) = \{c\}$. Thus, $(Z_1, Z_2, \ldots, Z_n, \{v_2\})$ is an $(n+1)$-homogeneous tuple. This completes the proof of the theorem. $\square$

3. Upper bounds

Erdős proved in [45] that for a natural number n, the set $P_n = \{ab : a, b \leq n\}$ has size $o(n^2)$. We base the proof of Theorem 1.3 on the observation that P_n is exactly the set of sizes of all induced subgraphs of a complete bipartite graph between two equal vertex classes of size n.

Let $H(x, y, z)$ be the number of natural numbers $n \leq x$ having a divisor in the interval $(y, z]$. Tenenbaum [105] showed that

$$H(x, y, z) = (1 + o(1))x \text{ if } \log y = o(\log z), z \leq \sqrt{x}. \tag{1}$$

Ford [56] proved that

$$H(x, y, 2y) = \Theta\left(\frac{x}{(\log y)^{\delta}(\log\log y)^{3/2}}\right) \text{ if } 3 \leq y \leq \sqrt{x}, \tag{2}$$

where $\delta = 1 - \frac{1+\log\log 2}{\log 2}$. Armed with these two facts, we can now prove Theorem 1.3.

Proof of Theorem 1.3. We shall take

$$A = \{k : \exists\, a, b \in \mathbb{N} \text{ with } k - 1 = ab \text{ and } \log k \leq a \leq b\}.$$

It follows from (1) that $H(x, \log x, \sqrt{x}) = (1 + o(1))x$; as an easy consequence, A has asymptotic density one. Now, for a fixed $k \in A$ with $k - 1 = ab$, consider a surjective k-colouring Δ of the complete graph on $\mathbb{N}$ which colours all the edges of the complete bipartite graph between $[a]$ and $[b + a] \setminus [a]$ with ab

distinct colours and all the other edges with the one colour not used so far. It is easy to then see that

$$\mathcal{F}_\Delta = \{a'b' + 1 : 1 \leq a' \leq a, 1 \leq b' \leq b\} \cup \{1\}.$$

Now, for any element $a'b' + 1 \in \mathcal{F}_\Delta$, $a/2^{i+1} < a' \leq a/2^i$ for some $i \geq 0$ and so $a'b' \leq ab/2^i$. Thus,

$$|\mathcal{F}_\Delta| \leq 1 + \sum_{i \geq 0} H\left(\frac{ab}{2^i}, \frac{a}{2^{i+1}}, \frac{a}{2^i}\right).$$

Using Ford's estimate (2) for $H(x, y, 2y)$ and the fact that $a \geq \log k$, we obtain that

$$\psi(k) = O\left(\frac{k}{(\log\log k)^\delta (\log\log\log k)^{3/2}}\right)$$

for all $k \in A$. □

4. Concluding remarks

We suspect that something much stronger than Corollary 2.2 is true.

Conjecture 4.1. *Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ be a k-colouring and suppose $n \geq 2$ is a natural number such that $k \geq \binom{n}{2} + 2$. Then $\mathcal{F}_\Delta \cap ([\binom{n+1}{2} + 1] \setminus [\binom{n}{2} + 1]) \neq \varnothing$.*

We return to this conjecture in the next chapter. There, we shall prove this conjecture and we shall also study some generalisations of this conjecture to uniform hypergraphs.

We strongly suspect that the function ψ studied in this chapter quite far from being monotone. We have shown that $\psi(\binom{n}{2} + 1) = n$ and $\psi(\binom{n+1}{2} + 1) = n + 1$, and it is an easy consequence of our results that $\psi(\binom{n}{2} + 2) = n + 1$. It would appear that even $\psi(\binom{n}{2} + 3)$ is much bigger than n.

Conjecture 4.2. *There is an absolute constant $c > 0$ such that $\psi(\binom{n}{2} + 3) > (1 + c)n$ for all natural numbers $n \geq 2$.*

The problem of determining ψ completely is of course still open. We cannot answer even the following question in its full generality.

PROBLEM 4.3. *Is* $\psi(k) = o(k)$ *for all* $k \in \mathbb{N}$*?*

CHAPTER 4

Approximations to infinite m-coloured complete hypergraphs

Joint work with Teeradej Kittipassorn.

1. Introduction

In the last chapter, we mentioned in passing that Stacey and Weidl, partially resolving a conjecture of Erickson, showed that there exists, for every fixed natural number $m > 2$ and all sufficiently large $k \in \mathbb{N}$, a k-colouring of the complete graph on $\mathbb{N}$ with no infinite m-coloured complete subgraph. In the light of this result, we study how close we can get to finding an infinite m-coloured complete subgraph (and more generally, an infinite m-coloured complete subhypergraph) in this chapter.

We briefly describe how the notations and definitions of the previous two chapters extend to the setting of hypergraphs. For a set X, we write $X^{(r)}$ for the family of all subsets of X of cardinality r; equivalently, $X^{(r)}$ is the complete r-uniform hypergraph on the vertex set X. By a *colouring* of a hypergraph, we mean a colouring of the edges of the hypergraph.

Let $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ be a surjective k-colouring of the edges of the complete r-uniform hypergraph on the natural numbers. As before, we say that a subset $X \subset \mathbb{N}$ is (*exactly*) *m-coloured* if $\Delta(X^{(r)})$, the set of values attained by Δ on the edges induced by X, has size exactly m. Let $\gamma_\Delta(X)$, or $\gamma(X)$ in short, denote the size of the set $\Delta(X^{(r)})$; in other words, every set X is $\gamma(X)$-coloured.

In this chapter, we shall study for fixed r and large k, the set of values m for which there exists an infinite m-coloured set with respect to a k-colouring

$\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$. As before, let us define

$$\mathcal{F}_\Delta = \{\gamma_\Delta(X) : X \subset \mathbb{N} \text{ such that } X \text{ is infinite}\}.$$

In the case of graphs, i.e., when $r = 2$, Stacey and Weidl [104], partially resolving a previously discussed conjecture of Erickson [52], showed using a probabilistic construction that for every $m > 2$, there is a constant C_m such that if $k > C_m$, then there is a k-colouring Δ of $\mathbb{N}^{(2)}$ such that $m \notin \mathcal{F}_\Delta$. Since an infinite m-coloured complete subgraph is not guaranteed to exist, we are naturally led to the question of whether we can find, given m, an infinite $\hat{m}$-coloured complete subgraph for some $\hat{m}$ close to m. In this chapter, we establish the following result.

Theorem 1.1. *For any k-colouring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ and any natural number $m \le k$, there exists an $\hat{m} \in \mathcal{F}_\Delta$ such that*

$$|m - \hat{m}| \le c_r m^{1-1/r} + O(m^{1-2/r})$$

where $c_r = r/(2(r!)^{1/r})$.

Theorem 1.1 is tight up to the $O(m^{1-2/r})$ term. To see this, let $k = \binom{n}{r} + 1$ for some $n \in \mathbb{N}$. We consider the 'small-rainbow colouring' Δ which colours all the edges induced by $[n]$ with $\binom{n}{r}$ distinct colours and all the remaining edges with the one colour that has not been used so far. In this case, we see that $\mathcal{F}_\Delta = \{\binom{i}{r} + 1 : i \le n\}$. Now let m be the positive integer closest to $(\binom{l}{r} + \binom{l+1}{r} + 2)/2$ for some natural number l such that $l < n$. It follows that $|m - \hat{m}| \ge \binom{l}{r-1}/2 - 1$ for each $\hat{m} \in \mathcal{F}_\Delta$; furthermore, it is easy to check that $\binom{l}{r-1}/2 = (c_r - o(1))m^{1-1/r}$.

In the case of graphs where $r = 2$, Theorem 1.1 tells us that for any finite colouring of the edges of the complete graph on $\mathbb{N}$ with m or more colours, there is an infinite $\hat{m}$-coloured complete subgraph for some $\hat{m}$ satisfying $|m - \hat{m}| \le \sqrt{m/2} + O(1)$; a careful analysis of the proof of Theorem 1.1 in this case allows us to replace the $O(1)$ term with an explicit constant, $1/2$.

We know from Theorem 1.1 that $\mathcal{F}_\Delta$ cannot contain very large gaps. Another natural question we are led to ask is if there are any sets, and in particular, intervals, that $\mathcal{F}_\Delta$ is guaranteed to intersect. More precisely, it was conjectured (see [91] and also, the previous chapter) that the small-rainbow colouring described above is extremal for graphs in the following sense.

CONJECTURE 1.2. *Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ be a k-colouring of the complete graph on $\mathbb{N}$ and suppose n is a natural number such that $k > \binom{n}{2} + 1$. Then $\mathcal{F}_\Delta \cap (\binom{n}{2} + 1, \binom{n+1}{2} + 1] \neq \varnothing$.*

In this chapter, we shall prove this conjecture. There are two natural generalisations of this conjecture to r-uniform hypergraphs which are equivalent to Conjecture 1.2 in the case of graphs.

The first comes from considering small-rainbow colourings; indeed we can ask whether $\mathcal{F}_\Delta \cap I_{r,n} \neq \varnothing$ when $k > \binom{n}{r} + 1$, where $I_{r,n}$ is the interval $(\binom{n}{r} + 1, \binom{n+1}{r} + 1]$.

The second comes from considering a different family of colourings which we call 'small-set colourings'. Let $k = \sum_{i=0}^{r} \binom{n}{i}$ and consider the surjective k-colouring Δ of $\mathbb{N}^{(r)}$ defined by $\Delta(e) = e \cap [n]$. Note that in this case, $\mathcal{F}_\Delta = \{\sum_{i=0}^{r} \binom{j}{i} : j \leq n\}$. Consequently, we can ask whether $\mathcal{F}_\Delta \cap J_{r,n} \neq \varnothing$ when $k > \sum_{i=0}^{r} \binom{n-1}{i}$, where $J_{r,n}$ is the interval $(\sum_{i=0}^{r} \binom{n-1}{i}, \sum_{i=0}^{r} \binom{n}{i}]$.

Note that both these questions are identical when $r = 2$. Indeed $\binom{n}{2} + \binom{n}{1} + \binom{n}{0} = \binom{n+1}{2} + 1$ and so $I_{2,n} = J_{2,n}$. When $r \geq 3$, we see that $J_{r,n}$ is longer than $I_{r,n}$; furthermore, for any fixed $r \geq 3$ and all sufficiently large n, we note that $J_{r,n}$ always intersects, and lies to the right of, $I_{r,n}$.

We shall demonstrate that the correct generalisation is the former. We shall first prove that the answer to the first question is in the affirmative, provided n is sufficiently large.

THEOREM 1.3. *For every $r \geq 2$, there exists a natural number $n_r \geq r - 1$ such that for any natural number $n \geq n_r$ and any k-colouring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ with $k > \binom{n}{r} + 1$, $\mathcal{F}_\Delta \cap I_{r,n} \neq \varnothing$.*

Using a result of Baranyai [21] on factorisations of uniform hypergraphs, we shall exhibit an infinite family of colourings that answer the second question negatively for every $r \geq 3$.

THEOREM 1.4. *For every $r \geq 3$, there exist infinitely many values of n for which there exists a k-colouring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ with $k > \sum_{i=0}^{r} \binom{n-1}{i}$ such that $\mathcal{F}_\Delta \cap J_{r,n} = \varnothing$.*

The rest of this chapter is organised as follows. In the next section, we shall prove Theorems 1.1, 1.3 and 1.4 and deduce Conjecture 1.2 from the proof of Theorem 1.3. We then conclude by mentioning some open problems.

2. Proofs of the main results

We start with the following lemma which we shall later use to prove both Theorems 1.1 and 1.3.

LEMMA 2.1. *Let $m \geq 2$ be an element of $\mathcal{F}_\Delta$. Then there exists a natural number $a = a(m, \Delta)$ such that*

(1) $\sum_{i=0}^{r} \binom{a}{i} \geq m$, *and*
(2) $\mathcal{F}_\Delta \cap [m - \min(\sum_{i=0}^{r-1} \binom{a-1}{i}, r(m-1)/a), m) \neq \varnothing$.

Futhermore, if

$$m = \sum_{i=t+1}^{r} \binom{a}{i} + s + 1$$

for some $s \geq 0$ and $0 \leq t+1 \leq r$, then

$$\mathcal{F}_\Delta \cap \left[\sum_{i=t+1}^{r} \binom{a-1}{i} + \left(1 - \frac{t}{a}\right)s + 1, m \right) \neq \varnothing.$$

PROOF. We start by establishing the following claim.

CLAIM 2.2. *There is an infinite m-coloured set $X \subset \mathbb{N}$ with a finite subset $A \subset X$ such that*

(1) the colour of every edge of X is determined by its intersection with A, i.e., if $e_1 \cap A = e_2 \cap A$, then $\Delta(e_1) = \Delta(e_2)$, and

(2) $\gamma(X \setminus \{v\}) < m$ for all $v \in A$.

Proof. To see this, let $W \subset \mathbb{N}$ be an infinite m-coloured set. For each colour $c \in \Delta(W^{(r)})$, pick an edge e_c in W of colour c and let $A = \bigcup_c e_c$ be the set of vertices incident to these edges. So $A \subset W$ is a finite m-coloured set. Let $A_1, A_2, \ldots, A_l$ be an enumeration of the subsets of A of size at most r. Note that this is the complete list of possible intersections of an edge with A. We now define a descending sequence of infinite sets $B_0 \supset B_1 \supset \cdots \supset B_l$ as follows. Let $B_0 = W \setminus A$. Having defined the infinite set B_{i-1}, we induce a colouring of the $(r - |A_i|)$-tuples T of B_{i-1}, by giving T the colour of the edge $A_i \cup T$. By Ramsey's Theorem, there is an infinite monochromatic subset $B_i \subset B_{i-1}$ with respect to this induced colouring, and so the edges of $A \cup B_i$ whose intersection with A is A_i have the same colour.

Hence, $X = A \cup B_l$ is an infinite m-coloured set satisfying property (1). Now, if we have a vertex $v \in A$ such that $\gamma(X \setminus \{v\}) = m$, we delete v from A. We repeat this until we are left with an m-coloured set X satisfying (1) and (2). $\square$

Let X and A be as guaranteed by Claim 2.2. Note that A is nonempty since $m \geq 2$. We shall prove the lemma with $a(m, \Delta) = |A|$. From the structure of X and A, we note that $\sum_{i=0}^{r} \binom{a}{i} \geq m$. That

$$\mathcal{F}_\Delta \cap \left[m - \min\left(\sum_{i=0}^{r-1} \binom{a-1}{i}, \frac{r(m-1)}{a} \right), m \right) \neq \varnothing$$

is a consequence of the following claim.

Claim 2.3. *There exist infinite sets $X_1, X_2 \subset X$ for which we have $m - \sum_{i=0}^{r-1} \binom{a-1}{i} \leq \gamma(X_1) < m$ and $m - r(m-1)/a \leq \gamma(X_2) < m$.*

Proof. Let $X_1 = X \setminus \{v\}$ for any $v \in A$. We know from Claim 2.2 that $\gamma(X_1) < m$. We shall now prove that $\gamma(X_1) \geq m - \sum_{i=0}^{r-1} \binom{a-1}{i}$; that is, the number of colours lost by removing v from X is at most $\sum_{i=0}^{r-1} \binom{a-1}{i}$. Since the colour of an edge is determined by its intersection with A, the number of colours lost is at most the numbers of subsets of A containing v of size at most r, which is precisely $\sum_{i=0}^{r-1}\binom{a-1}{i}$.

Next, we shall prove that there is a subset $X_2 \subset X$ such that $m - r(m-1)/a \leq \gamma(X_2) < m$. Let $A = \{v_1, v_2, \ldots, v_a\}$ and let

$$C_i = \Delta\big(X^{(r)}\big) \setminus \Delta\big((X \setminus \{v_i\})^{(r)}\big)$$

be the set of colours lost by removing v_i from X; since $\gamma(X \setminus \{v_i\}) < m$ for all $v_i \in A$, it follows that $C_i \neq \varnothing$. For each colour $c \in \Delta(X^{(r)})$, pick an edge e_c of colour c, and let $A_c = e_c \cap A$; in particular, we take $A_{c_\varnothing} = \varnothing$, where $c_\varnothing$ is the colour corresponding to an empty intersection with A. Since every edge of colour $c \in C_i$ contains v_i, we double count the number of times a colour is counted in the sum $\sum_{i=1}^{a}|C_i|$ to obtain

$$\sum_{i=1}^{a}|C_i| \leq \sum_{c \neq c_\varnothing}|A_c| \leq r(m-1),$$

and so there exists an i such that $0 < |C_i| \leq r(m-1)/a$; the claim follows by taking $X_2 = X \setminus \{v_i\}$. $\square$

We finish the proof of the lemma by establishing the following claim.

Claim 2.4. *If we can write $m = \sum_{i=t+1}^{r}\binom{a}{i} + s + 1$, then*

$$\mathcal{F}_\Delta \cap \left[\sum_{i=t+1}^{r}\binom{a-1}{i} + \left(1 - \frac{t}{a}\right)s + 1, m\right) \neq \varnothing.$$

Proof. As in the proof of Claim 2.3, for each colour $c \in \Delta(X^{(r)})$, pick an edge e_c of colour c, and let $A_c = e_c \cap A$; in particular, let $A_{c_\varnothing} = \varnothing$. We know from Claim 2.2 that edges of X of distinct colours cannot have the same intersection with A. Consequently, all the A_c are distinct subsets of A, each of size at most r. Hence,

$$\sum_{c \neq c_\varnothing}|A_c| \leq \sum_{i=t+1}^{r} i\binom{a}{i} + ts.$$

Arguing as in the proof of Claim 2.3, we conclude that there exists a vertex $v \in A$ such that the number of colours lost by removing v from X is at most $(\sum_{i=t+1}^{r} i\binom{a}{i} + ts)/a$. Therefore

$$\begin{aligned}\gamma(X \setminus \{v\}) &\geq m - \frac{1}{a}\left(\sum_{i=t+1}^{r} i\binom{a}{i} + ts\right)\\ &= m - \left(\sum_{i=t+1}^{r}\binom{a-1}{i-1} + \frac{ts}{a}\right)\\ &= \sum_{i=t+1}^{r}\binom{a-1}{i} + \left(1-\frac{t}{a}\right)s + 1,\end{aligned}$$

and so

$$\mathcal{F}_\Delta \cap \left[\sum_{i=t+1}^{r}\binom{a-1}{i} + \left(1-\frac{t}{a}\right)s+1, m\right) \neq \varnothing. \qquad \square$$

The lemma follows from Claims 2.2, 2.3 and 2.4. We are done. $\square$

Having established Lemma 2.1, it is easy to deduce both Theorem 1.1 and 1.3 from the lemma.

Proof of Theorem 1.1. Let $t = m + c_r m^{1-1/r}$. We may assume that $m > r^r/r!$ since otherwise $m = O(1)$ and there is nothing to prove. Also, if $t \geq k$, then the result follows easily by taking $\hat{m} = k$ so we may assume that $t < k$. Let $\hat{t}$ be the smallest element of $\mathcal{F}_\Delta$ greater than t. Applying Lemma 2.1 to $\hat{t}$, we find an $\hat{m} \in \mathcal{F}_\Delta$ such that $\hat{m} \leq t$ and

$$\hat{m} \geq \hat{t} - \min\left(\sum_{i=0}^{r-1}\binom{a-1}{i}, \frac{r(\hat{t}-1)}{a}\right)$$

for some natural number a. Now if $a \geq (r!m)^{1/r} > r$, then

$$\hat{m} \geq \hat{t} - \frac{r(\hat{t}-1)}{a} \geq \hat{t}\left(1-\frac{r}{a}\right) \geq t\left(1-\frac{r}{a}\right)$$

and so it follows that $\hat{m} \geq m - c_r m^{1-1/r} - O(m^{1-2/r})$. If $a < (r!m)^{1/r}$ on the other hand, then using the fact that

$$\begin{aligned}\hat{m} &\geq \hat{t} - \sum_{i=0}^{r-1} \binom{a-1}{i} \\ &\geq t - \frac{a^{r-1}}{(r-1)!} - O(a^{r-2}) \\ &\geq t - \frac{(r!m)^{1-1/r}}{(r-1)!} - O(m^{1-2/r}),\end{aligned}$$

it follows once again that $\hat{m} \geq m - c_r m^{1-1/r} - O(m^{1-2/r})$. □

Proof of Theorem 1.3. If $k \leq \binom{n+1}{r} + 1$, we are done since $k \in \mathcal{F}_\Delta$. So suppose that $k > \binom{n+1}{r} + 1$. Let m be the smallest element of $\mathcal{F}_\Delta$ such that $m > \binom{n+1}{r} + 1$; hence, $\mathcal{F}_\Delta \cap (\binom{n+1}{r} + 1, m) = \varnothing$. Now, since $m \geq 2$, there exists by Lemma 2.1, a natural number a such that

$$\mathcal{F}_\Delta \cap \left[m - \frac{r(m-1)}{a}, \binom{n+1}{r} + 1\right] \neq \varnothing.$$

To prove the theorem, it is sufficient to show that $m - r(m-1)/a > \binom{n}{r} + 1$. We know from Lemma 2.1 that $\sum_{i=0}^{r} \binom{a}{i} \geq m > \binom{n+1}{r} + 1$. If n is sufficiently large, we must have $a \geq n$. Indeed, if $a \leq n-1$, then

$$\begin{aligned}\binom{n+1}{r} + 1 - \sum_{i=0}^{r} \binom{a}{i} &\geq \binom{n+1}{r} - \binom{n-1}{r} - \binom{n-1}{r-1} - \sum_{i=1}^{r-2} \binom{n-1}{i} \\ &= \binom{n}{r-1} - \sum_{i=1}^{r-2} \binom{n-1}{i} > 0,\end{aligned}$$

where the last inequality holds for all sufficiently large n since coefficient of the highest power of n in the expression is positive; this is a contradiction.

If $a \geq n+1$, then

$$\begin{aligned}m - \frac{r(m-1)}{a} &= (m-1)\Big(1 - \frac{r}{a}\Big) + 1 \\ &> \binom{n+1}{r}\Big(1 - \frac{r}{n+1}\Big) + 1\end{aligned}$$

$$= \binom{n}{r} + 1$$

since $m > \binom{n+1}{r} + 1$ and $n \geq r - 1$.

We now deal with the case $a = n$. First, we write $m = \binom{n}{r} + \binom{n}{r-1} + s + 1$. Since $m > \binom{n+1}{r} + 1$ and $\binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r}$, we see that $s > 0$. By Lemma 2.1, it follows that

$$\mathcal{F}_\Delta \cap \left[\binom{n}{r} + \left(1 - \frac{r-2}{n}\right) s + 1, m \right) \neq \varnothing.$$

Since $n \geq r - 1$ and $s > 0$, the result follows. $\square$

A careful inspection of the proof of Theorem 1.3 shows that when $r = 2$, the statement holds for all $n \in \mathbb{N}$. Indeed, we only need n to be large enough to ensure that $\binom{n}{r-1} - \sum_{i=1}^{r-2} \binom{n-1}{i} > 0$; when $r = 2$, this holds for all $n \in \mathbb{N}$. We hence obtain a proof of Conjecture 1.2.

We now turn to the proof of Theorem 1.4. We require a result of Baranyai [21] which states that the set of edges of the complete r-uniform hypergraph on l vertices can be partitioned into perfect matchings when $r \mid l$.

Proof of Theorem 1.4. We shall show that if n is sufficiently large and $(r-1) \mid (n+1)$, then there is a surjective k-colouring Δ of $\mathbb{N}^{(r)}$ with $k > \sum_{i=0}^{r} \binom{n-1}{i}$ and $\mathcal{F}_\Delta \cap J_{r,n} = \varnothing$. We shall define a colouring of $\mathbb{N}^{(r)}$ such that the colour of an edge e is determined by its intersection with a set A of size $n+1$, say $A = [n+1]$. Let $\mathcal{B}$ be the family of all subsets of A of size at most r. For $B \in \mathcal{B}$, we denote the colour assigned to all the edges e such that $e \cap A = B$ by c_B.

To define our colouring, we shall construct a partition $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ with $\varnothing \in \mathcal{B}_2$. Then for every $B \in \mathcal{B}_2$, we set c_B to be equal to $c_\varnothing$. Finally, we take the colours c_B for $B \in \mathcal{B}_1$ to all be distinct and different from $c_\varnothing$. Hence, the number of colours used is $k = |\mathcal{B}_1| + 1$. It remains to construct this partition of $\mathcal{B}$.

Since $(r-1) \mid (n+1)$, by Baranyai's theorem there exists an ordering

$$B_1, B_2, \ldots, B_{\binom{n+1}{r-1}}$$

of the subsets of A of size $r-1$ such that for all $0 \le t \le \binom{n}{r-2}$, the family

$$\left\{ B_{\left(\frac{n+1}{r-1}\right)t+1}, B_{\left(\frac{n+1}{r-1}\right)t+2}, \ldots, B_{\left(\frac{n+1}{r-1}\right)(t+1)} \right\}$$

is a perfect matching. Let $\mathcal{B}_1 = \{B_1, B_2, \ldots, B_s\} \cup \{B \in \mathcal{B} : |B| = r\}$, where

$$s = \sum_{i=0}^{r} \binom{n}{i} - \binom{n+1}{r} = \sum_{i=0}^{r-2} \binom{n}{i}.$$

Our colouring is well defined because $0 \le s \le \binom{n+1}{r-1}$ for all sufficiently large n; here, the second inequality follows immediately by considering the coefficient of the highest power of n. Observe that

$$k = |\mathcal{B}_1| + 1 = \binom{n+1}{r} + s + 1 = \sum_{i=0}^{r} \binom{n}{i} + 1.$$

We shall show that the second largest element of $\mathcal{F}_\Delta$ is at most $\sum_{i=0}^{r} \binom{n-1}{i}$. Note that any $X \subset \mathbb{N}$ with $\gamma(X) < k$ cannot contain A. As before, let C_i be the set of colours lost by removing $i \in A$ from $\mathbb{N}$, i.e.,

$$C_i = \Delta\big(\mathbb{N}^{(r)}\big) \setminus \Delta\big((\mathbb{N} \setminus \{i\})^{(r)}\big).$$

We shall complete the proof by showing that $k - |C_i| \le \sum_{i=0}^{r} \binom{n-1}{i}$ for all $i \in A$.

Note that our construction ensures that $||C_i| - |C_j|| \le 1$ for all $i, j \in A$. Now, observe that

$$\sum_{i=1}^{n+1} |C_i| = \sum_{B \in \mathcal{B}_1} |B| = r\binom{n+1}{r} + (r-1)s,$$

and so $|C_i| \ge (r\binom{n+1}{r} + (r-1)s)/(n+1) - 1$ for all $i \in A$. Hence,

$$k - |C_i| \le \left(\binom{n+1}{r} + s + 1 \right) - \frac{1}{n+1}\left(r\binom{n+1}{r} + (r-1)s \right) + 1$$

$$
\begin{aligned}
&= \binom{n}{r} + \left(1 - \frac{r-1}{n+1}\right) s + 2 \\
&= \binom{n}{r} + \left(1 - \frac{r-1}{n+1}\right)\left(\sum_{i=0}^{r} \binom{n}{i} - \binom{n+1}{r}\right) + 2 \\
&\le \sum_{i=0}^{r} \binom{n-1}{i},
\end{aligned}
$$

where the last inequality holds when $r \ge 4$ for all sufficiently large n. To see this, we note that verifying the last inequality reduces to checking that

$$\left(1 - \frac{r-1}{n+1}\right)\left(\sum_{i=0}^{r-2} \binom{n}{i}\right) + 2 \le \sum_{i=0}^{r-2} \binom{n-1}{i}.$$

To check this, we first note that for each $0 \le i \le r-3$, we have

$$\left(1 - \frac{r-1}{n+1}\right)\binom{n}{i} = \frac{n(n+2-r)}{(n+1)(n-i)}\binom{n-1}{i} \le \binom{n-1}{i}.$$

and that furthermore, we have

$$\binom{n-1}{r-2} - \left(1 - \frac{r-1}{n+1}\right)\binom{n}{r-2} = \frac{1}{n+1}\binom{n-1}{r-2} > 2$$

for all sufficiently large n since $r \ge 4$.

When $r = 3$, it is easy to check that $s = n+1$ and so s is divisible by $(n+1)/(r-1) = (n+1)/2$. Consequently, in this case, $|C_i| = |C_j|$ for $i, j \in A$. Hence,

$$
\begin{aligned}
k - |C_i| &\le \left(\binom{n+1}{3} + s + 1\right) - \frac{1}{n+1}\left(r\binom{n+1}{3} + 2s\right) \\
&= \binom{n}{3} + \left(1 - \frac{2}{n+1}\right)(n+1) + 1 \\
&= \sum_{i=0}^{3} \binom{n-1}{i}.
\end{aligned}
$$

This completes the proof. $\square$

3. Concluding remarks

We conclude by mentioning two open problems. We proved that for any k-colouring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ and every sufficiently large natural number n, $\mathcal{F}_\Delta \cap I_{r,n} \neq \varnothing$ provided $k > \binom{n}{r} + 1$. A careful analysis of our proof shows that the result holds when $n \geq (5/2 + o(1))r$; we chose not to give details to keep the presentation simple. However, we suspect that the result should hold as long as $n \geq r - 1$ but a proof eludes us.

To state the next problem, let us define

$$\psi_r(k) = \min_{\Delta:\mathbb{N}^{(r)} \twoheadrightarrow [k]} |\mathcal{F}_\Delta|.$$

A consequence of Theorem 1.3 is that $\psi_r(k) \geq (1 - o(1))(r!k)^{1/r}$. Turning to the question of upper bounds for ψ_r, the small-rainbow colouring shows that the lower bound that we get from Theorem 1.3 is tight infinitely often, i.e., when k is of the form $\binom{n}{r} + 1$ for some $n \in \mathbb{N}$. However, when k is not of this form, the obvious generalisations of the small-rainbow colouring fail to give us good upper bounds for $\psi_r(k)$. We saw in the previous chapter that

$$\psi_2(k) = O\left(\frac{k}{(\log\log k)^\delta (\log\log\log k)^{3/2}}\right)$$

for almost all natural numbers k and some absolute constant $\delta > 0$. For every $r \geq 2$, by colouring a copy of $\mathbb{N}^{(2)}$ in $\mathbb{N}^{(r)}$ (corresponding to all the r-element subsets of $\mathbb{N}$ containing some fixed $(r-2)$-element set) as in the previous chapter, and all the other edges of $\mathbb{N}^{(r)}$ with a different colour, we see that $\psi_r(k) = o(k)$ almost all natural numbers k. It would be very interesting to decide if, in fact, $\psi_r(k) = o(k)$ for all $k \in \mathbb{N}$.

Part 2

Probabilistic combinatorics

CHAPTER 5

Transference for the Erdős–Ko–Rado theorem, I

Joint work with Béla Bollobás and Andrei Raigorodskii.

1. Introduction

In this chapter, our aim is to investigate the stability of a central result in extremal set theory due to Erdős, Ko and Rado [46] about uniform intersecting families of sets. A family of sets $\mathcal{A}$ is said to be *intersecting* if $A \cap B \neq \varnothing$ for all $A, B \in \mathcal{A}$. We are interested in intersecting families where all the sets have the same size; writing $[n]$ for the set $\{1, 2, \ldots, n\}$ and $[n]^{(r)}$ for the family of all the subsets of $[n]$ of cardinality r, the Erdős–Ko–Rado theorem asserts that if $\mathcal{A} \subset [n]^{(r)}$ is intersecting and $n \geq 2r$, then $|\mathcal{A}| \leq \binom{n-1}{r-1}$ and that equality is only achieved, if $n > 2r$, when $\mathcal{A}$ is a *star*; for $x \in [n]$, the *star centred at* x is the family of all the r-element subsets of $[n]$ containing x. Extending this result, Hilton and Milner [68] determined, when $n > 2r$, the largest size of a uniform intersecting family not contained entirely in a star. Many extensions of the Erdős–Ko–Rado theorem and the Hilton–Milner theorem have since been proved; furthermore, very general stability results about the structure of intersecting families have been proved by Friedgut [58], Dinur and Friedgut [43], and Keevash and Mubayi [74].

Here, we shall investigate a different notion of stability and prove a 'sparse random' analogue of the Erdős–Ko–Rado theorem which strengthens the Erdős–Ko–Rado theorem significantly when r is small compared to n.

To translate the Erdős–Ko–Rado theorem to the random setting, it will be helpful to reformulate the theorem as a statement about Kneser graphs. For natural numbers $n, r \in \mathbb{N}$ with $n \geq r$, the Kneser graph $K(n, r)$ is the graph whose vertex set is $[n]^{(r)}$ where two r-element sets $A, B \in [n]^{(r)}$ are adjacent if and only if $A \cap B = \varnothing$. Observe that a family $\mathcal{A} \subset [n]^{(r)}$ is an intersecting family if and only if $\mathcal{A}$ is an independent set in $K(n, r)$. Writing $\alpha(G)$ for the size of the largest independent set in a graph G, the Erdős–Ko–Rado theorem asserts that $\alpha(K(n, r)) = \binom{n-1}{r-1}$ when $n \geq 2r$; furthermore, when $n > 2r$, the only independent sets of this size are stars.

Let us now randomly delete the edges of the Kneser graph $K(n, r)$ retaining them with some probability p, independently of each other. When is the independence number of this random subgraph equal to $\binom{n-1}{r-1}$? It turns out that when r is much smaller than n, an analogue of the Erdős–Ko–Rado theorem continues to be true even after we delete practically all the edges of the Kneser graph!

This kind of phenomenon, namely the validity of classical extremal results for surprisingly sparse random structures, has received a lot of attention over the past twenty five years.

Perhaps the first result of this kind in extremal graph theory was proved by Babai, Simonovits, and Spencer [8] who showed that an analogue of Mantel's Theorem is true for certain random graphs. Mantel's Theorem states that the largest triangle free subgraph and the largest bipartite subgraph of K_n, the complete graph on n vertices, have the same size. Babai, Simonovits, and Spencer proved that the same holds for the Erdős-Rényi random graph $G(n, p)$ with high probability when $p \geq 1/2 - \delta$ for some absolute constant $\delta > 0$. In other words, they show that Mantel's theorem is 'stable' in the sense that it holds not only for the complete graph but that it holds *exactly* for random subgraphs of the complete graph as well. Improving upon results of Brightwell, Panagiotou and Steger [34], DeMarco and Kahn [41] have recently shown that this phenomenon continues to hold even when the random graph $G(n, p)$ is very sparse; they show in particular that it suffices to take $p \geq C(\log n/n)^{1/2}$

for some absolute constant $C > 0$, and that this is best possible up to the value of the absolute constant.

The first such transference results in Ramsey theory were proved by Rödl and Ruciński [97, 98] and there have been many related Ramsey theoretic results since; see, for example, [59, 99, 79].

Phenomena of this kind have also been observed in additive combinatorics. Roth's theorem [101], a central result in additive combinatorics, states that for every $\delta > 0$ and all sufficiently large n, every subset of $[n] = \{1, 2, \dots, n\}$ of density δ contains a three-term arithmetic progression. Kohayakawa, Rödl and Łuczak [78] proved a random analogue, showing that such a statement holds not only for $[n]$ but also, with high probability, for random subsets of $[n]$ of density at least $Cn^{-1/2}$, where $C > 0$ is an absolute constant.

Another classical result in additive combinatorics, due to Diananda and Yap [42], is that the largest sum-free subset of $\mathbb{Z}_{2n}$ is the set of odd numbers. Balogh, Morris and Samotij [20] proved that the same is true of random subsets of $\mathbb{Z}_{2n}$ of density at least $(1+\varepsilon)(\log n/3n)^{1/2}$ with high probability (for any fixed $\varepsilon > 0$ and n sufficiently large), and also that this no longer the case when the density is less than $(1-\varepsilon)(\log n/3n)^{1/2}$. Thus, there is a *sharp threshold* at $(\log n/3n)^{1/2}$ for the stability of this extremal result; an extension of this sharp threshold result to all even-order Abelian groups has recently been proved by Bushaw, Collares Neto, Morris and Smith [35].

Perhaps the most striking application of such transference principles in additive combinatorics is the Green–Tao theorem [66] on primes in arithmetic progressions.

These results constitute a tiny sample of the large number of beautiful results which have been proved in this setting. Very general transference theorems have been proved by Conlon and Gowers [40] and Schacht [103], and more recently, by Balogh, Morris and Samotij [19] and Saxton and Thomason [102]. We refer the interested reader to the surveys of Łuczak [87] and Rödl and Schacht [100] for a more detailed account of such results.

Returning to the question at hand, our aim in this chapter, as we remarked before, is to investigate the independence number of random subgraphs of $K(n,r)$; for related work on the independence number of random *induced* subgraphs of $K(n,r)$, see the paper of Balogh, Bohman and Mubayi [13]. Let $K_p(n,r)$ denote the random subgraph of $K(n,r)$ obtained by retaining each edge of $K(n,r)$ independently with probability p. The main question of interest is the following.

Problem 1.1. *For what $p > 0$ is $\alpha(K_p(n,r)) = \binom{n-1}{r-1}$ with high probability?*

For constant r and n sufficiently large, a partial answer was provided by Bogolyubskiy, Gusev, Pyaderkin and Raigorodskii [25, 24]: they studied random subgraphs of $K(n,r,s)$, where $K(n,r,s)$ is the graph whose vertex set is $[n]^{(r)}$ where two r-element sets $A, B \in [n]^{(r)}$ are adjacent if and only if $|A \cap B| = s$; in the case $s = 0$ (which corresponds to the Kneser graph), they established that $\alpha(K_{1/2}(n,r)) = (1 + o(1))\binom{n-1}{r-1}$ with high probability.

We shall do much more and answer Question 1.1 exactly when r is small compared to n (more precisely, when $r = o(n^{1/3})$). To state our result, it will be convenient to define the threshold function

$$p_c(n,r) = \frac{(r+1)\log n - r\log r}{\binom{n-1}{r-1}}. \tag{3}$$

As we shall see, this is the threshold density at which one expects to find a vertex in $K_p(n,r)$ which has no edges to a maximal independent set of the original Kneser graph $K(n,r)$. With this definition in place, we can now state our main result.

Theorem 1.2. *Fix a real number $\varepsilon > 0$ and let $r = r(n)$ be a natural number such that $2 \le r(n) = o(n^{1/3})$. Then as $n \to \infty$,*

$$\mathbb{P}\left(\alpha(K_p(n,r)) = \binom{n-1}{r-1}\right) \to \begin{cases} 1 & \text{if } p \ge (1+\varepsilon)p_c(n,r) \\ 0 & \text{if } p \le (1-\varepsilon)p_c(n,r). \end{cases}$$

Furthermore, when $p \geq (1+\varepsilon)p_c$, with high probability, the only independent sets of size $\binom{n-1}{r-1}$ in $K_p(n,r)$ are the trivial ones, namely, stars.

The rest of this chapter is organised as follows. We establish some notation and collect together some standard facts in Section 2. Most of the work involved in proving Theorem 1.2 is in establishing the upper bound on the critical density; we do this in Section 3. We complete the proof of Theorem 1.2 by proving a matching lower bound in Section 4. We conclude with some discussion in Section 5.

2. Preliminaries

2.1. Notation. Given $x \in [n]$ and $\mathcal{A} \subset [n]^{(r)}$, we write $\mathcal{S}_x$ for the star centred at x, and $\mathcal{A}_x$ for the subfamily of $\mathcal{A}$ consisting of those sets (of $\mathcal{A}$) that contain x, i.e., $\mathcal{A}_x = \mathcal{A} \cap \mathcal{S}_x$. The maximum degree $d(\mathcal{A})$ of a family $\mathcal{A} \subset [n]^{(r)}$ is defined to be the maximum cardinality, over all $x \in [n]$, of the subfamily $\mathcal{A}_x$, and we write $e(\mathcal{A})$ for the number of edges induced by $\mathcal{A}$ in $K(n,r)$. Since any pair of intersecting sets $A, B \in \mathcal{A}$ both belong to at least one subfamily $\mathcal{A}_x$, we get the following estimate for $e(\mathcal{A})$ which is useful when the maximum degree of $\mathcal{A}$ is small.

Proposition 2.1. *For any $\mathcal{A} \subset [n]^{(r)}$,*

$$e(\mathcal{A}) \geq \binom{|\mathcal{A}|}{2} - \sum_{x \in [n]} \binom{|\mathcal{A}_x|}{2}. \qquad \square$$

To ease the notational burden, in the rest of this chapter, we shall write $\mathbf{V} = \binom{n}{r}$ for the size of $[n]^{(r)}$, and $\mathbf{N} = \binom{n-1}{r-1}$ for the size of a star. Also, given $x \in [n]$ and a set $A \in [n]^{(r)}$ not containing x, we shall write $\mathbf{M} = \binom{n-r-1}{r-1}$ for the number of sets of $\mathcal{S}_x$ disjoint from A.

A word on asymptotic notation; we use the standard $o(1)$ notation to denote any function that tends to zero as n tends to infinity. Here and elsewhere, the variable tending to infinity will always be n unless we explicitly specify otherwise.

2.2. Estimates. Next, we collect some standard estimates that we shall use repeatedly; for ease of reference, we list them as propositions below. We refer the reader to Chapter 1 of [27] for the proofs of these claims.

Let us start with a weak form of Stirling's approximation for the factorial function.

PROPOSITION 2.2. *For all $n \in \mathbb{N}$,*

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \leq n! \leq e^{1/12n}\sqrt{2\pi n}\left(\frac{n}{e}\right)^n. \qquad \square$$

In fact, the following crude bounds for the binomial coefficients will often be sufficient for our purposes.

PROPOSITION 2.3. *For all $n, r \in \mathbb{N}$,*

$$\left(\frac{n}{r}\right)^r \leq \binom{n}{r} \leq \frac{n^r}{r!} \leq \left(\frac{en}{r}\right)^r. \qquad \square$$

Also, we will need the following standard inequality concerning the exponential function.

PROPOSITION 2.4. *For every $x \in \mathbb{R}$ such that $|x| \leq 1/2$,*

$$e^{x-x^2} \leq 1 + x \leq e^x. \qquad \square$$

Although our last proposition is also very simple, we prove it here for the sake of completeness. Recall that $\mathbf{N} = \binom{n-1}{r-1}$ and $\mathbf{M} = \binom{n-r-1}{r-1}$.

PROPOSITION 2.5. *If $r = r(n) = o(n^{1/2})$, then $\mathbf{N} - \mathbf{M} = o(\mathbf{N})$. Furthermore, if $r = o(n^{1/3})$, then $\mathbf{N} - \mathbf{M} = o(\mathbf{N}/r)$.*

PROOF. Both claims follow from the observation that

$$\begin{aligned}\mathbf{N} - \mathbf{M} &= \binom{n-1}{r-1} - \binom{n-r-1}{r-1} \\ &= \sum_{i=1}^{r} \binom{n-i}{r-1} - \binom{n-i-1}{r-1}\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{r} \binom{n-i-1}{r-2} \\
&\leq r\binom{n-2}{r-2} = \frac{r(r-1)}{n-1}\mathbf{N}. \qquad \square
\end{aligned}
$$

3. Upper bound for the critical threshold

We now turn to our proof of Theorem 1.2. In this section, we shall bound the critical threshold from above, i.e., we shall prove that a random analogue of the Erdős–Ko–Rado theorem holds if $p > (1+\varepsilon)p_c(n,r)$ where $p_c(n,r)$ is given by (3).

Let us remind the reader before we begin that for us, a star in the Kneser graph is a maximal trivial intersecting family of sets (and this should not be confused with the graph-theoretic notion of a star).

Proof of the upper bound in Theorem 1.2. Let $0 < \varepsilon < 1/2$ and set $p = p(n) = (1+\varepsilon)p_c(n,r)$. We shall prove that with high probability, the independence number of $K_p(n,r)$ is $\mathbf{N}$, and that furthermore, the only independent sets of size $\mathbf{N}$ in $K_p(n,r)$ are stars. Since we are working with monotone properties, it suffices to prove this result for ε small enough and so we lose nothing by assuming $0 < \varepsilon < 1/2$.

For each $i \geq 1$, let X_i be the number of families $\mathcal{A} \subset [n]^{(r)}$ inducing an independent set in $K_p(n,r)$ such that $|\mathcal{A}| = \mathbf{N}$ and $d(\mathcal{A}) = \mathbf{N} - i$. Also, let Y be the number of independent families $\mathcal{A} \subset [n]^{(r)}$ such that $|\mathcal{A}| = \mathbf{N}+1$ and $d(\mathcal{A}) = \mathbf{N}$; in other words, independent families of size $\mathbf{N}+1$ which contain an entire star.

Our aim is to show that with high probability, the random variables defined above are all equal to zero. This then implies the lower bound on the critical threshold; since every X_i is equal to zero, every independent set in $K_p(n,r)$ of cardinality at least $\mathbf{N}$ must contain an entire star, and since Y is also equal to zero, the only independent sets of cardinality at least $\mathbf{N}$ are stars.

We start by computing $\mathbb{E}[Y]$. We know that for any star $\mathcal{S}$, any $A \in [n]^{(r)} \backslash \mathcal{S}$ is disjoint from $\mathbf{M}$ elements of $\mathcal{S}$, and so,

$$\mathbb{E}[Y] = \binom{n}{1}\binom{\mathbf{V}-\mathbf{N}}{1}(1-p)^{\mathbf{M}}. \tag{4}$$

When $r = o(n^{1/3})$ (indeed, when $r = o(n^{1/2})$), we know from Proposition 2.5 that $\mathbf{M} = (1+o(1))\mathbf{N}$. Since $p = (1+\varepsilon)((r+1)\log n - r\log r)/\mathbf{N}$, we see that

$$\begin{aligned}
\mathbb{E}[Y] &\leq n\mathbf{V}(1-p)^{(1+o(1))\mathbf{N}} \\
&\leq n\left(\frac{en}{r}\right)^r \exp\left((-1+o(1))p\mathbf{N}\right) \\
&\leq n\left(\frac{en}{r}\right)^r \exp\left((1+\varepsilon+o(1))(r\log r - (r+1)\log n)\right) \\
&\leq \left(\frac{er}{n}\right)^{(\varepsilon+o(1))r} \leq n^{-(\varepsilon+o(1))2r/3} = o(1).
\end{aligned}$$

By Markov's inequality, we know that $\mathbb{P}(Y > 0) \leq \mathbb{E}[Y]$ and it follows that Y is zero with high probability.

We now turn our attention to the X_i. To keep our argument simple, we distinguish three cases: we first deal with small values of i where the X_i count families of very large maximum degree, then we consider families of large (but not huge) maximum degree, and in the final case, we deal with families of small maximum degree.

Case 1: Very large maximum degree. Unfortunately, when i is small, it is not true that $\mathbb{E}[X_i]$ goes to zero as n grows. For constant i, $\mathbb{E}[X_i] \geq n\binom{\mathbf{N}}{i}\binom{\mathbf{V}-\mathbf{N}}{i}(1-p)^{(i+o(1))\mathbf{N}}$. When $r = 3$ and $i = 2$ for example, it follows that

$$\begin{aligned}
\mathbb{E}[X_2] &\geq n\binom{\binom{n-1}{2}}{2}\binom{\binom{n}{3}-\binom{n-1}{2}}{2}(1-p)^{(2+o(1))\mathbf{N}} \\
&\geq n^{o(1)}\frac{n^{11}}{n^{8(1+\varepsilon)}} \geq n^{3-8\varepsilon+o(1)},
\end{aligned}$$

which grows with n when ε is small enough. However, if we compute $\mathrm{Var}\,[X_2]$, we are encouraged to find that $\mathrm{Var}\,[X_2]/\mathbb{E}[X_2]^2$ is bounded away from zero;

indeed, we observe similar behaviour for any fixed value of i and larger r as well. We therefore adopt a different strategy to bound $\mathbb{P}(X_i > 0)$ for small i.

For $j \geq i$, let $X_{i,j}$ be the number of families $\mathcal{A} \subset [n]^{(r)}$ inducing a *maximal* independent set in $K_p(n,r)$ such that $d(\mathcal{A}) = \mathbf{N} - i$ and $|\mathcal{A}| = \mathbf{N} + j - i$. If $X_i > 0$, then clearly $X_{i,j} > 0$ for some $j \geq i$. To compute $\mathbb{E}[X_{i,j}]$, we note that any family $\mathcal{A}$ counted by $X_{i,j}$ can be described by specifying a star $\mathcal{S}$, a subfamily $\mathcal{A}_1 \subset \mathcal{S}$ of i sets missing from $\mathcal{S}$, and another family $\mathcal{A}_2$ of cardinality j disjoint from $\mathcal{S}$ such that

(1) all the edges between $\mathcal{S} \setminus \mathcal{A}_1$ and $\mathcal{A}_2$ in $K(n,r)$ are absent in $K_p(n,r)$ (since $\mathcal{A}$ is independent), and
(2) each set in $\mathcal{A}_1$ is adjacent to at least one set in $\mathcal{A}_2$ in $K_p(n,r)$ (because $\mathcal{A}$ is a maximal independent set).

The number of edges between $\mathcal{S} \setminus \mathcal{A}_1$ and $\mathcal{A}_2$ is at least $j(\mathbf{M} - i)$ since any set not in a star is adjacent to precisely $\mathbf{M}$ sets in the star in $K(n,r)$. Also, the probability that a set in $\mathcal{A}_1$ has a neighbour in $\mathcal{A}_2$ in $K_p(n,r)$ is at most jp. Therefore, we have

$$\mathbb{E}[X_{i,j}] \leq n\binom{\mathbf{N}}{i}\binom{\mathbf{V}}{j}(1-p)^{j(\mathbf{M}-i)}(jp)^i.$$

We look at the ratio of the upper bounds for $\mathbb{E}[X_{i,j+1}]$ and $\mathbb{E}[X_{i,j}]$ above and note that this ratio is at most

$$\mathbf{V}(1-p)^{\mathbf{M}-i}(1+1/j)^i.$$

When $1 \leq i \leq \varepsilon\mathbf{N}/2$ and $j \geq i$, we see, using the fact that $\mathbf{M} = (1+o(1))\mathbf{N}$, that

$$\begin{aligned}\mathbf{V}(1-p)^{\mathbf{M}-i}(1+1/j)^i &\leq e\Big(\frac{en}{r}\Big)^r \exp\big(-(1+\varepsilon/2-\varepsilon^2/2+o(1))p_c(n,r)\mathbf{N}\big) \\ &\leq e^{r+1}\Big(\frac{r}{n}\Big)^{\varepsilon r/5} = o(1).\end{aligned}$$

Consequently, we have

$$\begin{aligned}
\mathbb{P}[X_i > 0] &\le \sum_{j\ge i} \mathbb{E}[X_{i,j}] \le 2n\binom{\mathbf{N}}{i}\binom{\mathbf{V}}{i}(1-p)^{i(\mathbf{M}-i)}(ip)^i \\
&\le 2n\left(\frac{e\mathbf{N}}{i}\right)^i\left(\frac{e\mathbf{V}}{i}\right)^i \exp\left(-i(1+\varepsilon/5)p_c(n,r)\mathbf{N}\right)\left(\frac{i(r+1)\log n}{\mathbf{N}}\right)^i \\
&\le 2e^{2i}n((r+1)\log n)^i\left(\frac{\mathbf{V}}{i}\right)^i \exp\left(-i(1+\varepsilon/5)p_c(n,r)\mathbf{N}\right) \\
&\le 2\left(\frac{e^{r+2}(r+1)\log n}{i}\left(\frac{r}{n}\right)^{\varepsilon r/5}\right)^i \le 2\left(e^{r+2}(r+1)\log n\left(\frac{r}{n}\right)^{\varepsilon r/5}\right)^i.
\end{aligned}$$

Summing this estimate for $i \le \varepsilon\mathbf{N}/2$, we get

$$\begin{aligned}
\sum_{i=1}^{\varepsilon\mathbf{N}/2}\mathbb{P}(X_i > 0) &\le \sum_{i=1}^{\varepsilon\mathbf{N}/2} 2\left(e^{r+2}(r+1)\log n\left(\frac{r}{n}\right)^{\varepsilon r/5}\right)^i \\
&\le 4\left(e^{r+2}(r+1)\log n\left(\frac{r}{n}\right)^{\varepsilon r/5}\right) = o(1),
\end{aligned}$$

and so by the union bound, with high probability, for each $1 \le i \le \varepsilon\mathbf{N}/2$, the random variable X_i is zero.

Case 2: Large maximum degree. Next, we consider the X_i with

$$\varepsilon\mathbf{N}/2 < i \le \mathbf{N}\left(1 - \frac{1-\varepsilon/2}{r+1}\right).$$

As noted earlier, for any star $\mathcal{S}$, the number of edges in $K(n,r)$ between a set $A \in [n]^{(r)} \setminus \mathcal{S}$ and a family $\mathcal{A} \subset \mathcal{S}$ is at least $|\mathcal{A}| - (\mathbf{N} - \mathbf{M})$. We know from Proposition 2.5 that $\mathbf{N} - \mathbf{M} = o(\mathbf{N}/r)$ when $r = o(n^{1/3})$; consequently, it follows that if $\mathcal{A} \subset [n]^{(r)}$ has cardinality $\mathbf{N}$ and $d(\mathcal{A}) \ge (1-\varepsilon/2)\mathbf{N}/(r+1)$, then $e(\mathcal{A}) \ge (1+o(1))d(\mathcal{A})(\mathbf{N} - d(\mathcal{A}))$.

To simplify calculations, let us define α by setting $i = \alpha\mathbf{N} = \alpha r\mathbf{V}/n$ where

$$\varepsilon/2 < \alpha \le (r+\varepsilon/2)/(r+1).$$

In this range, we see that

$$\mathbb{E}[X_i] \le n\binom{\mathbf{N}}{i}\binom{\mathbf{V}}{i}(1-p)^{(1+o(1))i(\mathbf{N}-i)}$$

$$\le n\Big(\frac{e}{\alpha}\Big)^{\alpha\mathbf{N}}\Big(\frac{en}{r\alpha}\Big)^{\alpha\mathbf{N}}\exp\big(-(1+\varepsilon+o(1))\alpha(1-\alpha)p_c(n,r)\mathbf{N}^2\big)$$
$$\le n\bigg(\frac{nr^{(1+\varepsilon+o(1))(1-\alpha)r}}{rn^{(1+\varepsilon+o(1))(1-\alpha)(r+1)}}\bigg)^{\alpha\mathbf{N}} \le n\Big(\frac{r}{n}\Big)^{(\varepsilon^2/4-\varepsilon^3/4+o(1))\mathbf{N}}.$$

The last two inequalities above are obtained by first collecting the $O(1)$ terms in the bound into the $o(1)$ terms in the exponent, and by then using the bounds on α. It follows that

$$\sum_{i=\varepsilon\mathbf{N}/2}^{\frac{(2r+\varepsilon)\mathbf{N}}{2(r+1)}}\mathbb{P}(X_i>0)\le n\mathbf{N}\Big(\frac{r}{n}\Big)^{(\varepsilon^2/4-\varepsilon^3/4+o(1))\mathbf{N}}=o(1),$$

and so with high probability, for each $\varepsilon\mathbf{N}/2<i\le(r+\varepsilon/2)\mathbf{N}/(r+1)$, the random variable X_i is zero.

Case 3: Small maximum degree. We shall complete the proof of the lower bound by showing that

$$\sum_{i>\frac{(2r+\varepsilon)\mathbf{N}}{2(r+1)}}\mathbb{E}[X_i]=o(1).$$

It turns out that in this range of i, somewhat surprisingly, it is significantly easier to deal with the case where r tends to infinity with n as opposed to the case where r is small.

Suppose first that $r\ge\log n$. Then

$$\frac{(1-\varepsilon/2)}{r+1}<\frac{1}{r+4}$$

for all large enough n. Observe that subgraph of $K(n,r)$ induced by a family $\mathcal{A}$ of cardinality $\mathbf{N}$ has minimum degree at least $\mathbf{N}-rd(\mathcal{A})$ and consequently, if $d(\mathcal{A})<\mathbf{N}/(r+4)$, then

$$e(\mathcal{A})\ge\frac{\mathbf{N}}{2}\bigg(\mathbf{N}-\frac{r\mathbf{N}}{r+4}\bigg)=\frac{2\mathbf{N}^2}{r+4}.$$

In this case, it follows that

$$\sum_{i > \frac{(2r+\varepsilon)\mathbf{N}}{2(r+1)}} \mathbb{E}[X_i] \le \binom{\mathbf{V}}{\mathbf{N}} (1-p)^{2\mathbf{N}^2/(r+4)}$$

$$\le \left(\frac{en}{r}\right)^{\mathbf{N}} \left(\frac{r}{n}\right)^{2r\mathbf{N}/(r+4)}$$

$$\le \left(\frac{er}{n}\right)^{(1+o(1))\mathbf{N}} = o(1)$$

which completes the proof when $r \ge \log n$.

Next, suppose that $r \le \log n$. When $r \le \log n$, it is not necessarily true (if $r = O(1)$ and ε is sufficiently small, for instance) that $(1-\varepsilon/2)/(r+1) < 1/(r+4)$. It turns out that in this case, we need a more careful estimate.

For a family $\mathcal{A} \subset [n]^{(r)}$ and each $x \in [n]$, define $\alpha_x = |\mathcal{A}_x|/\mathbf{N}$. Note that $\sum_{x=1}^n \alpha_x = r$. Recall that Proposition 2.1 tells us that

$$e(\mathcal{A}) \ge \binom{|\mathcal{A}|}{2} - \sum_{x\in[n]} \binom{|\mathcal{A}_x|}{2} \ge \binom{\mathbf{N}}{2}\left(1 - \sum_{x\in[n]} \alpha_x^2\right).$$

Let $\mathcal{A} \subset [n]^{(r)}$ be such that $|\mathcal{A}| = \mathbf{N}$ and $d(\mathcal{A}) < (1-\varepsilon/2)\mathbf{N}/(r+1)$. For such a family $\mathcal{A}$, let $D = D_{\mathcal{A}}$ be the set of $x \in [n]$ such that $\alpha_x \ge (\log n)^{-2}$. Since $\sum_{x=1}^n \alpha_x = r$, we see that $|D| \le r(\log n)^2 \le (\log n)^3$.

Lemma 3.1. *Fix $D = D_{\mathcal{A}}$ and the values of $|\mathcal{A}_x|$ for $x \in D$. Subject to these restrictions, the expected number of families $\mathcal{A} \subset [n]^{(r)}$ of maximum degree at most $(1-\varepsilon/2)\mathbf{N}/(r+1)$ which induce independent sets in $K_p(n,r)$ is at most $(r/n)^{(3/10+o(1))\mathbf{N}}$.*

Proof. Since $\sum_{x=1}^n \alpha_x = r$, it follows (by convexity, for example) that $\sum_{x\in[n]\setminus D} \alpha_x^2$ is at most $r(\log n)^{-2} \le (\log n)^{-1} = o(1)$. Consequently,

$$e(\mathcal{A}) \ge \frac{\mathbf{N}^2}{2}\left(1 + o(1) - \sum_{x\in[n]} \alpha_x^2\right) \ge \frac{\mathbf{N}^2}{2}\left(1 + o(1) - \sum_{x\in D} \alpha_x^2\right),$$

and so the probability that a family $\mathcal{A}$ as in the statement of the lemma induces an independent set is at most

$$
\begin{aligned}
(1-p)^{e(\mathcal{A})} &\leq \exp\left(-\frac{p\mathbf{N}^2}{2}\left(1+o(1)-\sum_{x\in D}\alpha_x^2\right)\right) \\
&\leq \left(\frac{r^{(1+o(1))r}}{n^{(1+o(1))(r+1)}} \prod_{x\in D}\left(\frac{n^{r+1}}{r^r}\right)^{\alpha_x^2}\right)^{\mathbf{N}/2}. \qquad (5)
\end{aligned}
$$

Next, we bound the number of ways in which we can choose $\mathcal{A}$ as in Lemma 3.1. Using the fact that $r \leq \log n$ and $|D| \leq (\log n)^3$, we first note that

$$
\begin{aligned}
\mathbf{N} \geq \left|\bigcup_{x\in D}\mathcal{A}_x\right| &\geq \sum_{x\in D}|\mathcal{A}_x| - \sum_{\substack{x,y\in D\\ x<y}}|\mathcal{A}_x\cap\mathcal{A}_y| \\
&\geq \sum_{x\in D}|\mathcal{A}_x| - |D|^2\binom{n-2}{r-2} \\
&\geq \left(\sum_{x\in D}\alpha_x - \frac{|D|^2 r}{n}\right)\mathbf{N} \geq \left(\sum_{x\in D}\alpha_x + o(1)\right)\mathbf{N}.
\end{aligned}
$$

It follows that

$$
\sum_{x\in D}\alpha_x \leq 1+o(1) < 1+1/10 \qquad (6)
$$

and

$$
|\mathcal{A}\setminus(\cup_{x\in D}\mathcal{A}_x)| < \mathbf{N}\left(1+1/5-\sum_{x\in D}\alpha_x\right).
$$

(Here, the choice of the constants 1/10 and 1/5 was arbitrary; any two sufficiently small constants would have sufficed.) Hence, the number of ways to choose $\mathcal{A}$ is at most

$$
\begin{aligned}
\binom{\mathbf{V}}{\mathbf{N}(6/5-\sum_{x\in D}\alpha_x)}\prod_{x\in D}\binom{\mathbf{N}}{\alpha_x\mathbf{N}} &\leq \left(\frac{10en}{r}\right)^{\mathbf{N}(6/5-\sum_{x\in D}\alpha_x)}\prod_{x\in D}\left(\frac{e}{\alpha_x}\right)^{\alpha_x\mathbf{N}} \\
&\leq 100^{\mathbf{N}}\left(\frac{n}{r}\right)^{6\mathbf{N}/5}\prod_{x\in D}\left(\frac{r}{\alpha_x n}\right)^{\alpha_x\mathbf{N}} \\
&\leq \left(\frac{n}{r}\right)^{(6/5+o(1))\mathbf{N}}\prod_{x\in D}\left(\frac{r}{\alpha_x n}\right)^{\alpha_x\mathbf{N}}. \qquad (7)
\end{aligned}
$$

From (5) and (7), we conclude that the expected number of independent families $\mathcal{A}$ as in the lemma is at most

$$\left(\frac{r^{(1+o(1))r/2-6/5}}{n^{(1+o(1))(r+1)/2-6/5}}\right)^{\mathbf{N}} \prod_{x\in D}\left(\left(\frac{r}{\alpha_x n}\right)\left(\frac{n^{r+1}}{r^r}\right)^{\alpha_x/2}\right)^{\alpha_x \mathbf{N}}$$

Now, note that$(r/\alpha n)(n^{r+1}/r^r)^{\alpha/2} < 1$ whenever $(\log n)^{-2} \le \alpha < (1-\varepsilon/2)/(r+1)$. Indeed, observe that the function f defined on the positive reals by

$$f(\alpha) = \frac{\alpha((r+1)\log n - r\log r)}{2} - \log\alpha + \log(r/n)$$

is convex; so to check that $f(\alpha) < 0$ when $(\log n)^{-2} \le \alpha \le (1-\varepsilon/2)/(r+1)$, it suffices to check that $f((\log n)^{-2}) < 0$ and $f((1-\varepsilon/2)/(r+1)) < 0$ and both conditions hold for all sufficiently large n when $r \le \log n$.

Therefore, we conclude that the expected number of independent families $\mathcal{A}$ as in the lemma is at most

$$\left(\frac{r^{(1+o(1))r/2-6/5}}{n^{(1+o(1))(r+1)/2-6/5}}\right)^{\mathbf{N}} \le \left(\frac{r^{(1+o(1))(r+1)/2-6/5}}{n^{(1+o(1))(r+1)/2-6/5}}\right)^{\mathbf{N}} \le \left(\frac{r}{n}\right)^{(3/10+o(1))\mathbf{N}},$$

where the last inequality above follows from the fact that $(r+1)/2-6/5 \ge 3/10$ for all $r \ge 2$. This completes the proof of Lemma 3.1. $\square$

Recall that if $r \le \log n$ and $d(\mathcal{A}) < (1-\varepsilon/2)\mathbf{N}/(r+1)$, then $|D_{\mathcal{A}}| \le (\log n)^3$. So the number of choices for the set $D_{\mathcal{A}}$ is clearly at most

$$\sum_{j=0}^{(\log n)^3}\binom{n}{j} \le (\log n)^3\binom{n}{(\log n)^3}. \tag{8}$$

We know from (6) that the values $|\mathcal{A}_x|$ for $x \in D_{\mathcal{A}}$ satisfy

$$\sum_{x\in D}|\mathcal{A}_x| \le 11\mathbf{N}/10$$

and so, the number of ways of selecting the values of $|\mathcal{A}_x|$ is at most

$$\binom{11\mathbf{N}/10 + (\log n)^3 + 1}{(\log n)^3} \le (2\mathbf{N})^{(\log n)^3}. \tag{9}$$

From Lemma 3.1, we conclude using (8) and (9) that

$$\sum_{i>\frac{(2r+\varepsilon)\mathbf{N}}{2(r+1)}} \mathbb{E}[X_i] \leq (\log n)^3 n^{(\log n)^3} (2\mathbf{N})^{(\log n)^3} \left(\frac{r}{n}\right)^{(3/10+o(1))\mathbf{N}}. \qquad (10)$$

It is easy to check that the right-hand side of (10) is $o(1)$ for every $2 \leq r \leq \log n$. Hence, with high probability, for each $i > (r+\varepsilon/2)\mathbf{N}/(r+1)$, the random variable X_i is zero; this completes the proof of the lower bound. □

4. Lower bound for the critical threshold

As in the previous section, let Y be the number of independent families in $K_p(n,r)$ of size $\mathbf{N}+1$ which contain an entire star.

Proof of the lower bound in Theorem 1.2. Turning to the lower bound, we shall assume that $p = (1-\varepsilon)p_c(n,r)$ for some fixed real number $\varepsilon > 0$ and we show using a simple second moment calculation that $Y > 0$ with high probability; consequently, the independence number of $K_p(n,r)$ is at least $\mathbf{N}+1$.

Recall (4) which says that

$$\mathbb{E}[Y] = \binom{n}{1}\binom{\mathbf{V}-\mathbf{N}}{1}(1-p)^{\mathbf{M}}.$$

Note that $\mathbf{N} = o(\mathbf{V})$ when $r = o(n^{1/3})$; it follows that

$$\begin{aligned}
\mathbb{E}[Y] &\geq (1+o(1))n\mathbf{V}(1-p)^{\mathbf{N}} \\
&\geq (1+o(1))\frac{n^{r+1}}{r!}\exp\left(-(p+p^2)\mathbf{N}\right) \\
&\geq \frac{n^{r+1}}{r!}\exp\left((1-\varepsilon+o(1))(r\log r-(r+1)\log n)\right) \\
&\geq \left(\frac{n}{r}\right)^{(\varepsilon+o(1))r},
\end{aligned}$$

and so $\mathbb{E}[Y] \to \infty$ when $p = (1-\varepsilon)p_c(n,r)$.

Therefore, to show that $Y > 0$ with high probability, it suffices to show that $\mathrm{Var}[Y] = o(\mathbb{E}[Y]^2)$ or equivalently, that $\mathbb{E}[(Y)_2] = (1+o(1))\mathbb{E}[Y]^2$, where $\mathbb{E}[(Y)_2] = \mathbb{E}[Y(Y-1)]$ is the second factorial moment of Y.

Note that

$$\mathbb{E}[(Y)_2] = \sum_{x,y,A,B} \mathbb{P}(\mathcal{S}_x \cup \{A\} \,\text{and}\, \mathcal{S}_y \cup \{B\} \,\text{are independent}),$$

the sum being over ordered 4-tuples (x, y, A, B) with $x, y \in [n]$, $A \in [n]^{(r)} \setminus \mathcal{S}_x$ and $B \in [n]^{(r)} \setminus \mathcal{S}_y$ such that $(x, A) \neq (y, B)$. Now, observe that

$$\begin{aligned}\sum_{x \neq y} \mathbb{P}(\mathcal{S}_x \cup \{A\} \,\text{and}\, \mathcal{S}_y \cup \{B\} \,\text{are independent}) &\leq (n^2)(\mathbf{V} - \mathbf{N})^2(1-p)^{(2-o(1))\mathbf{M}} \\ &= (1+o(1))\mathbb{E}[Y]^2,\end{aligned}$$

and

$$\begin{aligned}\sum_{x=y, A \neq B} \mathbb{P}(\mathcal{S}_x \cup \{A\} \,\text{and}\, \mathcal{S}_y \cup \{B\} \,\text{are independent}) &\leq n(\mathbf{V} - \mathbf{N})^2(1-p)^{2\mathbf{M}} \\ &= o(\mathbb{E}[Y]^2).\end{aligned}$$

By Chebyshev's inequality, we conclude that $Y > 0$ with high probability and so, the independence number of $K_p(n, r)$ is at least $\mathbf{N} + 1$. □

5. Concluding remarks

The condition $r = o(n^{1/3})$ in our results seems somewhat artificial; we would expect the same formula for the critical threshold to hold for much larger r as well. We suspect that this formula in fact gives the exact value of the critical threshold when $r = o(n)$ but we are unable to prove this presently.

The size of the critical window also merits study. Our proof (for large r) works even when we are a factor of $(1 + 6/r)$ away from the critical threshold; it is possible that the critical window is much smaller and it is an interesting problem to determine the size of the critical window precisely.

Of course, one would be interested to know what happens for larger r as well. When r/n is bounded away from $1/2$, we suspect it should be possible to demonstrate stability of the Erdős–Ko–Rado theorem at $p = 1/2$, say. We return to this question in the next chapter.

CHAPTER 6

Transference for the Erdős–Ko–Rado theorem, II

Joint work with József Balogh and Béla Bollobás.

1. Introduction

In this chapter, we shall continue the work begun in the last chapter and once again, our objective will be a transference result the Erdős–Ko–Rado theorem. Recall that for natural numbers $n, r \in \mathbb{N}$ with $n \geq r$, the Kneser graph $K(n, r)$ is the graph whose vertex set is $[n]^{(r)}$ where two r-element sets $A, B \in [n]^{(r)}$ are adjacent if and only if $A \cap B = \varnothing$. Observe that a family $\mathcal{A} \subset [n]^{(r)}$ is an intersecting family if and only if $\mathcal{A}$ induces an independent set in $K(n, r)$. Writing $\alpha(G)$ for the size of the largest independent set in a graph G, the Erdős–Ko–Rado theorem asserts that $\alpha(K(n, r)) = \binom{n-1}{r-1}$ when $n \geq 2r$ and furthermore, when $n > 2r$, the only independent sets of this size are stars.

Let $K_p(n, r)$ denote the random subgraph of $K(n, r)$ obtained by retaining each edge of $K(n, r)$ independently with probability p. In the previous chapter, we asked the following natural question: is $\alpha(K_p(n, r)) = \binom{n-1}{r-1}$? We proved, when $r = r(n) = o(n^{1/3})$, that the answer to this question is in the affirmative even after practically all the edges of the Kneser graph have been deleted. More precisely, we showed that in this range, there exists a (very small) critical probability $p_c(n, r)$ with the following property: as $n \to \infty$, if $p/p_c > 1$ then with high probability, $\alpha(K_p(n, r)) = \binom{n-1}{r-1}$ and the only independent sets of this size in $K_p(n, r)$ are stars, while if $p/p_c < 1$, then $\alpha(K_p(n, r)) > \binom{n-1}{r-1}$ with high probability. In this chapter, we shall prove a transference theorem that holds for much bigger values of $r = r(n)$; however, unlike in the previous chapter, we are unable to establish a sharp threshold. Our main result is the following.

Theorem 1.1. *For every $\varepsilon > 0$, there exist constants $c = c(\varepsilon) > 0$ and $c' = c'(\varepsilon) > 0$ with $c < c'$ such that for all $n, r \in \mathbb{N}$ with $r \leq (1/2 - \varepsilon)n$,*

$$\mathbb{P}\left(\alpha(K_p(n,r)) = \binom{n-1}{r-1}\right) \to \begin{cases} 1 & \text{if } p \geq \binom{n-1}{r-1}^{-c} \\ 0 & \text{if } p \leq \binom{n-1}{r-1}^{-c'} \end{cases}$$

as $n \to \infty$. In particular, with high probability, $\alpha(K_{1/2}(n,r)) = \binom{n-1}{r-1}$.

All the work in proving Theorem 1.1 is in showing that $c(\varepsilon)$ exists. The existence of $c'(\varepsilon)$, on the other hand, is an easy exercise. Indeed, a second moment calculation identical to that in Section 4 of Chapter 5 shows that if $p \leq \binom{n-r-1}{r-1}^{-1}$, then with high probability, there exists a star and an r-set not contained in the star all the edges between which are absent in $K_p(n,r)$. We can then check that if $r \leq (1/2 - \varepsilon)n$, then there exists a $c' = c'(\varepsilon)$ such that $\binom{n-r-1}{r-1} \geq \binom{n-1}{r-1}^{c'}$. Indeed, using Stirling's approximation, this reduces to verifying that $(n-r)\log(r/(n-r))/n\log(r/n)$ is uniformly bounded away from zero when $r \leq (1/2 - \varepsilon)n$; this is straightforward because the function decreases with r. Hence, if $p \leq \binom{n-1}{r-1}^{-c'}$, then $\alpha(K_p(n,r)) \geq \binom{n-1}{r-1} + 1$ with high probability.

Let us briefly describe some of the ideas that go into the proof of Theorem 1.1. We shall prove two results which, taken together, show that a large family $\mathcal{A} \subset [n]^{(r)}$ without a large intersecting subfamily must necessarily contain many pairs of disjoint sets, or in other words, must induce many edges in $K(n,r)$; we do this in Section 3. We put together the pieces and give the proof of Theorem 1.1 in Section 4. In Section 5, we briefly describe some approaches to improving the dependence of $c(\varepsilon)$ on ε in Theorem 1.1. We conclude with some discussion in Section 6.

2. Preliminaries

Henceforth, a 'family' will be a uniform family on $[n]$ unless we specify otherwise. To ease the notational burden, we adopt the following notational convention: when n and r are clear from the context, we write $\mathbf{V} = \binom{n}{r}$, $\mathbf{N} = \binom{n-1}{r-1}$, $\mathbf{M} = \binom{n-r-1}{r-1}$ and $\mathbf{R} = \binom{2r}{r}$.

We need a few results from extremal set theory, some classical and some more recent. The first result that we will need, due to Hilton and Milner [68], bounds the cardinality of a nontrivial uniform intersecting family. Writing $\mathcal{A}_x$ for the subfamily of a family $\mathcal{A}$ that consists of those sets containing x, we have the following.

Theorem 2.1. *Let $n, r \in \mathbb{N}$ and suppose that $n > 2r$. If $\mathcal{A} \subset [n]^{(r)}$ is an intersecting family with $|\mathcal{A}| \geq \mathbf{N} - \mathbf{M} + 2$, then there exists an $x \in [n]$ such that $\mathcal{A} = \mathcal{A}_x$.* □

The next result we shall require, due to Friedgut [58], is a quantitative extension of the Hilton–Milner theorem which says that any sufficiently large uniform intersecting family must resemble a star.

Theorem 2.2. *For every $\varepsilon > 0$, there exists a $C = C(\varepsilon) > 0$ such that for all $n, r \in \mathbb{N}$ with $\varepsilon n \leq r \leq (1/2 - \varepsilon)n$, the following holds: if $\mathcal{A} \subset [n]^{(r)}$ is an intersecting family and $|\mathcal{A}| = \mathbf{N} - k$, then there exists an $x \in [n]$ for which $|\mathcal{A}_x| \geq \mathbf{N} - Ck$.* □

We will also need the following well-known inequality for cross-intersecting families due to Bollobás [26].

Theorem 2.3. *Let $(A_1, B_1), \ldots, (A_m, B_m)$ be pairs of disjoint r-element sets such that $A_i \cap B_j \neq \varnothing$ for $i, j \in [m]$ whenever $i \neq j$. Then $m \leq \mathbf{R}$.* □

Finally, we shall require a theorem of Kruskal [80] and Katona [72]. For a family $\mathcal{A} \subset [n]^{(r)}$, its shadow in $[n]^{(k)}$, denoted $\partial^{(k)}\mathcal{A}$, is the family of those k-sets contained in some member of $\mathcal{A}$. For $x \in \mathbb{R}$ and $r \in \mathbb{N}$, we define the *generalised binomial coefficient* $\binom{x}{r}$ by setting

$$\binom{x}{r} = \frac{x(x-1)\ldots(x-r+1)}{r!}.$$

The following convenient formulation of the Kruskal–Katona theorem is due to Lovász [86].

THEOREM 2.4. *Let $n, r, k \in \mathbb{N}$ and suppose that $k \leq r \leq n$. If the cardinality of $\mathcal{A} \subset [n]^{(r)}$ is $\binom{x}{r}$ for some real number $x \geq r$, then $|\partial^{(k)}\mathcal{A}| \geq \binom{x}{k}$.* □

To avoid clutter, we omit floors and ceilings when they are not crucial. We use the standard $o(1)$ notation to denote any function that tends to zero as n tends to infinity; the variable tending to infinity will always be n unless we explicitly specify otherwise.

3. The number of disjoint pairs

Given a family $\mathcal{A}$, we write $e(\mathcal{A})$ for the number of disjoint pairs of sets in $\mathcal{A}$; equivalently, $e(\mathcal{A})$ is the number of edges in the subgraph of the Kneser graph induced by $\mathcal{A}$. In this section, we give some bounds for $e(\mathcal{A})$.

We denote by $\mathcal{A}^*$ the largest intersecting subfamily of a family $\mathcal{A}$; if this subfamily is not unique, we take any subfamily of maximum cardinality. We write $\ell(\mathcal{A}) = |\mathcal{A}| - |\mathcal{A}^*|$ for the difference between the cardinality of $\mathcal{A}$ and the largest intersecting subfamily of $\mathcal{A}$.

Trivially, we have $e(\mathcal{A}) \geq \ell(\mathcal{A})$. Our first lemma says that we can do much better than this trivial bound when $\ell(\mathcal{A})$ is large.

LEMMA 3.1. *Let $n, r \in \mathbb{N}$. For any $\mathcal{A} \subset [n]^{(r)}$,*

$$e(\mathcal{A}) \geq \frac{\ell(\mathcal{A})^2}{2\mathbf{R}}.$$

To prove this lemma, we need the notion of an induced matching. An *induced matching of size m* in a graph G is a set of $2m$ vertices inducing a subgraph consisting of m independent edges; equivalently, we refer to these m edges as an induced matching of size m. The *induced-matching number* of G, in notation, $m(G)$, is the maximal size of an induced matching in G.

Proposition 3.2. *Let $G = (V, E)$ be a graph with $m(G) = m \geq 1$. Then*

$$|E| \geq \frac{k^2}{4m}$$

where $k = |V| - \alpha(G)$.

Proof. Let us choose $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_m\}$ so that the edges $x_1y_1, \dots, x_my_m$ constitute an induced matching. Let $Z = \Gamma(X \cup Y)$ be the set of neighbours of the vertices in $X \cup Y$; thus $X \cup Y \subset Z$. Since $m(G) = m$, the set $V(G) \setminus Z$ is independent, and so $|Z| \geq k$. Since some vertex in $X \cup Y$ has at least $|Z|/2m$ neighbours, we conclude that $\Delta(G) \geq |Z|/2m \geq k/2m$ where $\Delta(G)$ is the maximum degree of G.

Now define a sequence of graphs $G = G_0 \supset G_1 \supset \cdots \supset G_k$ and a sequence of vertices $x_0, x_1, \dots, x_k$ by taking x_i to be a vertex of G_i of maximal degree and G_{i+1} to be the graph obtained from G_i by deleting x_i. We know from our earlier arguments that $\Delta(G_i) \geq (k-i)/2m$, and so $|E| \geq \sum_{i=0}^{k} \Delta(G_i) \geq k^2/4m$. □

To apply the previous proposition, we need the following corollary of Theorem 2.3 the proof of which is implicit in [18]; we include the short proof here for completeness.

Proposition 3.3. *For $n \geq 2r$, the induced-matching number of $K(n, r)$ is*

$$m(K(n,r)) = \binom{2r-1}{r-1} = \frac{\mathbf{R}}{2}.$$

Proof. Let $A_1B_1, \dots, A_mB_m$ be an induced matching in $K(n, r)$. For $m + 1 \leq i \leq 2m$, we set $A_i = B_{i-m}$ and $B_i = A_{i-m}$. We apply Theorem 2.3 to the pairs $(A_1, B_1), \dots, (A_{2m}, B_{2m})$ and conclude that $2m \leq \mathbf{R}$.

The $\mathbf{R}/2$ partitions of $[2r]$ into disjoint r-sets form an induced matching, and so $m(K(n, r)) = \mathbf{R}/2$, as claimed. □

Proof of Lemma 3.1. The lemma follows by applying Proposition 3.2 to $G_{\mathcal{A}}$, the subgraph of the Kneser graph $K(n, r)$ induced by $\mathcal{A}$. □

Note that Lemma 3.1 is only effective when $\ell(\mathcal{A}) \geq 2\mathbf{R}$. The next, somewhat technical, lemma complements Lemma 3.1 by giving a better bound when $\ell(\mathcal{A})$ is small provided the size of $\mathcal{A}$ is large.

LEMMA 3.4. *For every $\varepsilon, \eta > 0$, there exist constants $\delta = \delta(\varepsilon, \eta) > 0$ and $C = C(\varepsilon) > 0$ with the following property: for all $n, r \in \mathbb{N}$ with $\varepsilon n \leq r \leq (1/2 - \varepsilon)n$, we have*

$$e(\mathcal{A}) \geq \ell(\mathcal{A})^{1+\delta} - C\ell(\mathcal{A})$$

for any family $\mathcal{A} \subset [n]^{(r)}$ with $|\mathcal{A}| = \mathbf{N}$ and $\ell(\mathcal{A}) \leq \mathbf{N}^{1-\eta}$.

To clarify, the $C(\varepsilon)$ in the statement of the lemma above is the same as the $C(\varepsilon)$ guaranteed by Theorem 2.2.

PROOF OF LEMMA 3.4. First, let us note that since we always have $e(\mathcal{A}) \geq \ell(\mathcal{A})$, it suffices to prove the lemma under the assumption that n is sufficiently large; indeed, the lemma would then follow for all $n \in \mathbb{N}$ with an appropriately smaller value of δ.

Let $\ell = \ell(\mathcal{A})$. We start by observing that most of $\mathcal{A}$ must be contained in a star. Indeed, as before, let $\mathcal{A}^*$ denote the largest intersecting subfamily of $\mathcal{A}$; by definition, $|\mathcal{A}^*| = \mathbf{N} - \ell$. Since we have assumed that $\varepsilon n \leq r \leq (1/2 - \varepsilon)n$, we may assume, by Theorem 2.2, that $|\mathcal{A}^*_n| \geq \mathbf{N} - C\ell$ where $C = C(\varepsilon)$ is as guaranteed by Theorem 2.2. Hence, $|\mathcal{A}_n| \geq |\mathcal{A}^*_n| \geq \mathbf{N} - C\ell$.

We also know that $|\mathcal{A}_n| \leq |\mathcal{A}^*| \leq \mathbf{N} - \ell$; let $\mathcal{B}$ be a subset of $\mathcal{A} \setminus \mathcal{A}_n$ of cardinality exactly ℓ. We shall bound $e(\mathcal{A})$ by counting the number of edges between $\mathcal{B}$ and $\mathcal{A}_n$ in $K(n, r)$.

Let us define

$$\mathcal{A}' = \{A \setminus \{n\} : A \in \mathcal{A}_n\} \subset [n-1]^{(r-1)}$$

and

$$\mathcal{B}' = \{[n-1] \setminus B : B \in \mathcal{B}\} \subset [n-1]^{(n-r-1)}.$$

Clearly, to count the number of edges between $\mathcal{A}_n$ and $\mathcal{B}$ in $K(n, r)$, it suffices to count the number of pairs (A', B') in $\mathcal{A}' \times \mathcal{B}'$ with $A' \subset B'$. This quantity is obviously bounded below by the number of sets $A' \in \mathcal{A}'$ contained in at least one $B' \in \mathcal{B}'$.

Since $\mathcal{A}' \subset [n-1]^{(r-1)}$ and $|\mathcal{A}'| \geq \mathbf{N} - C\ell$, the number of sets $A' \in \mathcal{A}'$ contained in some $B' \in \mathcal{B}'$ is at least $|\partial^{(r-1)}\mathcal{B}'| - C\ell$. Consequently,

$$e(\mathcal{A}) \geq |\partial^{(r-1)}\mathcal{B}'| - C\ell.$$

We shall show that there exists a $\delta = \delta(\varepsilon, \eta) > 0$ such that, under the conditions of the lemma, $|\partial^{(r-1)}\mathcal{B}'| \geq \ell^{1+\delta}$.

We deduce the existence of such a δ from Theorem 2.4, the Kruskal–Katona theorem. We may assume that

$$\ell = |\mathcal{B}'| = \binom{x}{n-r-1}$$

for some real number $x \geq n - r - 1$. It follows from Theorem 2.4 that

$$|\partial^{(r-1)}\mathcal{B}'| \geq \binom{x}{r-1}.$$

Let us put $r = (1/2 - \beta)n$ and $x = \vartheta n$. We now calculate what values β and ϑ can take. We know that $\varepsilon \leq \beta \leq 1/2 - \varepsilon$. Since $x \geq n - r - 1$, we also know that $\vartheta \geq 1/2 + \beta - 1/n \geq 1/2 + \beta/2$. On the other hand, since

$$\binom{\vartheta n}{(1/2+\beta)n} = \ell \leq \mathbf{N}^{1-\eta} = \binom{n-1}{r-1}^{1-\eta} \leq \binom{n}{r}^{1-\eta} = \binom{n}{(1/2-\beta)n}^{1-\eta},$$

it follows from Stirling's approximation for the factorial function that there exists some $\delta'(\varepsilon, \eta) > 0$ such that $\vartheta \leq 1 - \delta'$.

Hence it suffices to check that there exists a $\delta = \delta(\varepsilon, \eta) > 0$ for which the inequality

$$\binom{\vartheta n}{(1/2-\beta)n - 1} \geq \binom{\vartheta n}{(1/2+\beta)n - 1}^{1+\delta}$$

holds for all $\beta \in [\varepsilon, 1/2 - \varepsilon]$ and $\vartheta \in [1/2 + \beta/2, 1 - \delta']$ as long as n is sufficiently large. This is easily checked using Stirling's formula. Indeed, let

$H(x) = -x\log_2(x) - (1-x)\log_2(1-x)$ be the binary entropy function. We know from Stirling's approximation that

$$\binom{\vartheta n}{(1/2-\beta)n-1} = \exp\left((\vartheta + o(1))H\left(\frac{1/2-\beta}{\vartheta}\right)n\right),$$

and similarly that

$$\binom{\vartheta n}{(1/2+\beta)n-1} = \exp\left((\vartheta + o(1))H\left(\frac{1/2+\beta}{\vartheta}\right)n\right).$$

Hence, it suffices to show that there exists a $\delta = \delta(\varepsilon, \delta')$ such that $H((1/2-\beta)/\vartheta) > (1+\delta)H((1/2+\beta)/\vartheta)$ for all $\beta \in [\varepsilon, 1/2-\varepsilon]$ and $\vartheta \in [1/2+\beta/2, 1-\delta']$. This follows easily from the fact that H is a concave function attaining its maximum at $1/2$ and the fact that $H(x) = H(1-x)$. □

4. Proof of the main result

Armed with Lemmas 3.1 and 3.4, we are now ready to prove Theorem 1.1.

Proof of Theorem 1.1. Let us fix $\varepsilon > 0$ and assume that $r \leq (1/2-\varepsilon)n$. Clearly, it is enough to prove Theorem 1.1 for all sufficiently small ε; it will be convenient to assume that $\varepsilon < 1/10$. As mentioned earlier, Bollobás, Narayanan and Raigorodskii have proved Theorem 1.1 in a much stronger form when $r = o(n^{1/3})$. So to avoid having to distinguish too many cases, we shall assume that r grows with n; for concreteness, let us suppose that $r \geq n^{1/4}$. A consequence of these assumptions is that in this range, $\mathbf{V}$, $\mathbf{N}$ and $\mathbf{M}$ all grow much faster than any polynomial in n.

Recall that $K_p(n,r)$ is the random subgraph of the Kneser graph $K(n,r)$ where we retain each edge of $K(n,r)$ independently with probability p. For each $\ell \geq 1$, let X_ℓ denote the (random) number of independent sets $\mathcal{A} \subset [n]^{(r)}$ in $K_p(n,r)$ with $|\mathcal{A}| = \mathbf{N}$ and $\ell(\mathcal{A}) = \ell$.

To prove Theorem 1.1, it clearly suffices to show that for some $c = c(\varepsilon) > 0$, all of the X_ℓ are zero with high probability provided $p \geq \mathbf{N}^{-c}$. We shall prove this by distinguishing three cases depending on which of Theorem 2.1, Lemma 3.1 and Lemma 3.4 is to be used.

Let $C = C(\varepsilon)$ be as in Theorem 2.2. Note that since $r \leq (1/2 - \varepsilon)n$, it is easy to check using Stirling's approximation that we can choose positive constants $c_m = c_m(\varepsilon)$ and $c_r = c_r(\varepsilon)$ such that $\mathbf{M} \geq \mathbf{N}^{c_m}$ and $\mathbf{R} \leq \mathbf{N}^{1-c_r}$.

We now set $L_m = \mathbf{N}^{c_m}/2$ and $L_r = \mathbf{N}^{1-c_r/4}$ and distinguish the following three cases.

Case 1: $\ell \leq L_m$. Let $\mathcal{A} \subset [n]^{(r)}$ be a family of cardinality $\mathbf{N}$ with $\ell(\mathcal{A}) = \ell$. Since

$$\ell \leq L_m = \mathbf{N}^{c_m}/2 \leq \mathbf{M} - 2,$$

we see that $\mathcal{A}^*$, the largest intersecting subfamily of $\mathcal{A}$, satisfies

$$|\mathcal{A}^*| = \mathbf{N} - \ell \geq \mathbf{N} - \mathbf{M} + 2.$$

It follows from Theorem 2.1 that there is an $x \in [n]$ for which $\mathcal{A}^*$ is contained in the star centred at x. Consider the ℓ sets in $\mathcal{A} \setminus \mathcal{A}^*$. Any such set is disjoint from exactly $\mathbf{M}$ members of the star centred at x and hence from at least $\mathbf{M} - \ell$ members of $\mathcal{A}^*$. This tells us that $e(\mathcal{A}) \geq \ell(\mathbf{M} - \ell)$. Since $\ell \leq \mathbf{M}/2$, it follows that

$$\begin{aligned}
\mathbb{E}[X_\ell] &\leq n\binom{\mathbf{N}}{\ell}\binom{\mathbf{V}}{\ell}(1-p)^{\ell(\mathbf{M}-\ell)} \\
&\leq n\binom{2^n}{\ell}^2 \exp(-p\ell\mathbf{M}/2) \\
&\leq \exp(2n\ell - p\ell\mathbf{M}/2).
\end{aligned}$$

Hence, if $c \leq c_m/2$ so that $p \geq \mathbf{N}^{-c_m/2}$, it is clear that

$$\sum_{\ell=1}^{L_m} \mathbb{E}[X_\ell] \leq \sum_{\ell=1}^{L_m} \exp\left(2n\ell - \frac{\ell\mathbf{N}^{c_m/2}}{2}\right) = o(1).$$

So with high probability, for each $1 \leq \ell \leq L_m$, the random variable X_ℓ is zero.

Case 2: $\ell \geq L_r$**.** Again, let $\mathcal{A} \subset [n]^{(r)}$ be a family of cardinality $\mathbf{N}$ with $\ell(\mathcal{A}) = \ell$. We know from Lemma 3.1 that

$$e(\mathcal{A}) \geq \frac{\ell^2}{2\mathbf{R}} \geq \frac{\mathbf{N}^{2-c_r/2}}{2\mathbf{N}^{1-c_r}} = \frac{\mathbf{N}^{1+c_r/2}}{2}.$$

So it follows that

$$\sum_{l \geq L_r} \mathbb{E}[X_\ell] \leq \binom{\mathbf{V}}{\mathbf{N}} \exp\left(-p\frac{\mathbf{N}^{1+c_r/2}}{2}\right) \leq \exp\left(n\mathbf{N} - p\frac{\mathbf{N}^{1+c_r/2}}{2}\right).$$

Hence if $c \leq c_r/4$ so that $p \geq \mathbf{N}^{-c_r/4}$, we have

$$\sum_{l \geq L_r} \mathbb{E}[X_\ell] \leq \exp\left(n\mathbf{N} - \frac{\mathbf{N}^{1+c_r/4}}{2}\right) = o(1).$$

So once again, with high probability, the sum $\sum_{\ell \geq L_r} X_\ell$ is zero.

Before we proceed further, let us first show that that we may now assume without loss of generality that $r \geq \varepsilon n$. This is because one can check that the arguments in Cases 1 and 2 together prove Theorem 1.1 when $r \leq \varepsilon n$ for all sufficiently small ε. It is easy to check using Stirling's formula that if ε is sufficiently small, indeed if $\varepsilon < 1/10$ for example, then it is possible to choose positive constants $c'_m(\varepsilon)$ and $c'_r(\varepsilon)$ so that for all $r \leq \varepsilon n$, we have $\mathbf{M} \geq \mathbf{N}^{c'_m}$, $\mathbf{R} \leq \mathbf{N}^{1-c'_r}$ and $\mathbf{N}^{c'_m}/2 \geq \mathbf{N}^{1-c'_r/4}$. So the arguments above yield a proof of Theorem 1.1 when $r \leq \varepsilon n$. Therefore, in the following, we assume that $r \geq \varepsilon n$.

Case 3: $L_m \leq \ell \leq L_r$**.** As before, consider any family $\mathcal{A} \subset [n]^{(r)}$ of cardinality $\mathbf{N}$ with $\ell(\mathcal{A}) = \ell$. First note that since $\varepsilon n \leq r \leq (1/2 - \varepsilon n)$ and $\ell \leq L_r = \mathbf{N}^{1-c_r/4}$ where c_r is a constant depending only on ε, by Lemma 3.4, there exists a $\delta = \delta(\varepsilon)$ such that

$$e(\mathcal{A}) \geq \ell^{1+\delta} - C\ell.$$

Since $\ell \geq L_m = \mathbf{N}^{c_m}/2$, it follows that

$$e(\mathcal{A}) \geq \ell^{1+\delta} - C\ell \geq \ell^{1+\delta/2}$$

for all sufficiently large n.

Next, consider $\mathcal{A}^*$, the largest intersecting subfamily of $\mathcal{A}$, of cardinality $\mathbf{N} - \ell$. We know from Theorem 2.2 that there exists an $x \in [n]$ such that $|\mathcal{A}^*_x| \geq \mathbf{N} - C\ell$ and so $|\mathcal{A}_x| \geq \mathbf{N} - C\ell$. It is then easy to see that

$$\begin{aligned}\mathbb{E}[X_\ell] &\leq n\binom{\mathbf{N}}{C\ell}\binom{\mathbf{V}}{C\ell}(1-p)^{\ell^{1+\delta/2}} \\ &\leq \exp\big(\ell\big(2Cn - p\ell^{\delta/2}\big)\big).\end{aligned}$$

Hence, if $c \leq c_m\delta/4$ so that $p \geq \mathbf{N}^{-c_m\delta/4}$, it follows that

$$\sum_{\ell=L_m}^{L_r} \mathbb{E}[X_\ell] \leq \sum_{\ell=L_m}^{L_r} \exp\big(\ell\big(2Cn - \mathbf{N}^{c_m\delta/4}/2\big)\big) = o(1)$$

and so with high probability, for each $L_m \leq \ell \leq L_r$, the random variable X_ℓ is zero.

Putting the different parts of our argument together, we find that if $0 < \varepsilon < 1/10$,

$$c = c(\varepsilon) = \min\left(\frac{c_m(\varepsilon)}{2}, \frac{c'_m(\varepsilon)}{2}, \frac{c_r(\varepsilon)}{4}, \frac{c'_r(\varepsilon)}{4}, \frac{c_m(\varepsilon)\delta(\varepsilon)}{2}\right)$$

and $p \geq \mathbf{N}^{-c}$, then for all $r = r(n) \leq (1/2 - \varepsilon)n$, we have

$$\mathbb{P}\left(\alpha(K_p(n,r)) = \binom{n-1}{r-1}\right) \to 1$$

as $n \to \infty$. This completes the proof of Theorem 1.1. □

5. Avenues for improvement

We briefly discuss how one might tighten up the arguments in Theorem 1.1 so as to improve the dependence of $c(\varepsilon)$ on ε in the result. However, since it seems unlikely to us that these methods will be sufficient to determine the precise critical threshold at which Theorem 1.1 ceases to hold, we shall keep the discussion in this section largely informal.

5.1. Containers for sparse sets in the Kneser graph. The first approach we sketch involves using ideas from the theory of 'graph containers' to count large sparse sets in the Kneser graph more efficiently.

The theory of graph containers was originally developed to efficiently count the number of independent sets in a graph satisfying some kind of 'supersaturation' condition. The basic principle used to construct containers for graphs can be traced back to the work of Kleitman and Winston [77]. A great deal of work has since gone into refining and generalising their ideas, culminating in the results of Balogh, Morris and Samotij [19] and Saxton and Thomason [102]; these papers also give a detailed account of the history behind these ideas and we refer the interested reader there for details about how the general methodology was developed. Here we shall content ourselves with a brief discussion of how these ideas might be used to improve the dependence of $c(\varepsilon)$ on ε in Theorem 1.1.

Let us write $Y_m = Y_m(n, r)$ for the number of families $\mathcal{A} \subset [n]^{(r)}$ with $|\mathcal{A}| = \mathbf{N}$ and $e(\mathcal{A}) = m$. Clearly, to show that $\alpha(K_p(n, r)) = \mathbf{N}$ with high probability, it suffices to show that $\sum_{m \geq 1} Y_m(1 - p)^m = o(1)$. Hence, it would be useful to have good estimates for Y_m. We shall derive some bounds for Y_m; see Theorem 5.2 below. These bounds are not strong enough (especially for small values of m) to prove Theorem 1.1. However, note that in our proof of Theorem 1.1, we use the somewhat cavalier bound of $\binom{\mathbf{V}}{\mathbf{N}}$ for the number of families $\mathcal{A}$ of size $\mathbf{N}$ for which $\ell(\mathcal{A})$ is equal to some prescribed value (in Case 2 of the proof); we can instead use Theorem 5.2 to count more efficiently.

To prove an effective container theorem, one needs to first establish a suitable supersaturation property. Lovász [85] determined the second largest eigenvalue of the Kneser graph; by combining Lovász's result with the expander mixing lemma, Balogh, Das, Delcout, Liu and Sharifzadeh [18] (see also [62]) proved the following supersaturation theorem for the Kneser graph.

Proposition 5.1. *Let $n, r, k \in \mathbb{N}$ and suppose that $n > 2r$ and $k \leq \mathbf{V} - \mathbf{N}$. If $\mathcal{A} \subset [n]^{(r)}$ has cardinality $\mathbf{N} + k$, then $e(\mathcal{A}) \geq k\mathbf{M}/2$.* □

Using Proposition 5.1, we prove the following container theorem for the Kneser graph.

Theorem 5.2. *For every $\varepsilon > 0$, there exists a $\hat{C} = \hat{C}(\varepsilon) > 0$ such that for every $\beta > 0$ and all $n, r, m \in \mathbb{N}$ with $\varepsilon n \le r \le (1/2 - \varepsilon)n$ and $m \le \mathbf{V}\binom{n-r}{r}/2$, the following holds: writing*

$$k_1 = \hat{C}\left(\frac{\mathbf{N}}{\beta\mathbf{M}} + \left(\frac{m\mathbf{N}}{\beta\mathbf{M}}\right)^{1/2}\right)$$

and

$$k_2 = k_1 + \hat{C}\beta\mathbf{N},$$

there exist, for $1 \le i \le \sum_{j=0}^{k_1}\binom{\mathbf{V}}{j}$, families $\mathcal{B}_i \subset [n]^{(r)}$ each of cardinality at most $\mathbf{N} + k_2$ with the property that each $\mathcal{A} \subset [n]^{(r)}$ with $e(\mathcal{A}) \le m$ is contained in one of these families.

The advantage of this formulation of Theorem 5.2 in terms of k_1, k_2 and β is that we can apply the theorem with a value of $\beta > 0$ suitably chosen for the application at hand.

It is easy to check from Theorem 5.2 that $Y_m = Y_m(n,r)$, the number of families $\mathcal{A} \subset [n]^{(r)}$ with $|\mathcal{A}| = \mathbf{N}$ and $e(\mathcal{A}) = m$, satisfies

$$\begin{aligned} Y_m(n,r) &\le \left(\sum_{j=0}^{k_1}\binom{\mathbf{V}}{j}\right)\binom{\mathbf{N}+k_2}{\mathbf{N}} = 2\binom{\mathbf{V}}{k_1}\binom{\mathbf{N}+k_2}{k_2} \le 2\binom{\mathbf{V}}{k_1}\binom{\mathbf{V}}{k_2} \\ &\le 2\exp\left(\hat{C}n\left(\beta\mathbf{N} + \frac{2\mathbf{N}}{\beta\mathbf{M}} + \left(\frac{4m\mathbf{N}}{\beta\mathbf{M}}\right)^{1/2}\right)\right) \end{aligned}$$

for all $\beta > 0$ such that $k_1 \le \mathbf{V}/2$. (Indeed, there are $\sum_{j=0}^{k_1}\binom{\mathbf{V}}{j}$ ways of choosing a container $\mathcal{B}$, and there are at most $\binom{\mathbf{N}+k_2}{\mathbf{N}}$ sets of size $\mathbf{N}$ in $\mathcal{B}$.) We can then optimise this bound by choosing β depending on how large m is in comparison to $\mathbf{M}$ and $\mathbf{N}$.

Proof of Theorem 5.2. We start by proving a lemma whose proof is loosely based on the methods of Saxton and Thomason [102]. Before we state the lemma, let us have some notation. Given a graph $G = (V, E)$ and

$U \subset V(G)$, we write

$$\mu(U) = \frac{|E(G[U])|}{|V|};$$

in other words, $\mu(U)$ is the number of edges induced by U divided by the number of vertices of G. Also, we write $\mathcal{P}(X)$ for the collection of all subsets of a set X.

LEMMA 5.3. *Let $G = (V, E)$ be a graph with average degree d and maximum degree Δ. For every $a \geq 0$ and $b > 0$, there is a map $\mathcal{C} : \mathcal{P}(V) \to \mathcal{P}(V)$ with the following property: for every $U \subset V$ with $\mu(U) \leq a$, there is a subset $T \subset V$ such that*

(1) $T \subset U \subset \mathcal{C}(T)$,
(2) $|T| \leq 2|V|(a/bd)^{1/2} + |V|/bd$, and
(3) $\mu(\mathcal{C}(T)) \leq 2\Delta(a/bd)^{1/2} + \Delta/bd + bd$.

PROOF. We shall describe an algorithm that constructs T given U. The algorithm will also construct $\mathcal{C}(T)$ in parallel; it will be clear from the algorithm that $\mathcal{C}(T)$ is entirely determined by T and in no way depends on U.

Fix a linear ordering of the vertex set V of G. If u and v are adjacent and u precedes v in our ordering, we call v a *forward neighbour* of u and u a *backward neighbour* of v. For a vertex $v \in V$, we write $F(v)$ for the set of its forward neighbours.

We begin by setting $T = \varnothing$ and $A = V$. We shall iterate through V in the order we have fixed and add vertices to T and remove vertices from A as we go along; at any stage, we write $\Gamma(T)$ to denote the set of those vertices which, at that stage, have k or more backward neighbours in T where k is the least integer strictly greater than $(abd)^{1/2}$.

As we iterate through the vertices of V in order, we do the following when considering a vertex v.

(1) If $v \in \Gamma(T)$, we remove v from A; if it is also the case that $v \in U$, then we add v to T.

(2) If $v \notin \Gamma(T)$, we consider the size of $S = F(v) \setminus \Gamma(T)$.

(a) If $|S| \geq bd$, we remove v from A; if it is also the case that $v \in U$, then we add v to T.

(b) If $|S| < bd$, we do nothing.

The algorithm outputs T and A when it terminates; we then set $\mathcal{C}(T) = A \cup T$. It is clear from the algorithm that $\mathcal{C}(T)$ is uniquely determined by T and that $T \subset U \subset \mathcal{C}(T)$. To see this, note that the decision to remove a vertex v from A depends only on which vertices preceding v belong to T. Therefore, we can reconstruct A, and hence $\mathcal{C}(T)$, using only the set T without any knowledge of the set U.

We first show that $|T| \leq 2|V|(a/bd)^{1/2} + |V|/bd$. Consider the partition $T = T_1 \cup T_2$ where T_1 consists of those vertices which were added to T on account of condition (1) and T_2 of those vertices which were added to T when considering condition (2a). The upper bound for $|T|$ follows from the following two claims.

Claim 5.4. *$|T_1| \leq |E(G[U])|/k$.*

Proof. Clearly, each vertex of T_1 has at least k backward neighbours in $T \subset U$. Hence, $k|T_1| \leq |E(G[U])|$. □

Claim 5.5. *$|T_2| \leq k|V|/bd$.*

Proof. Let us mark all the edges from v to $F(v) \setminus \Gamma(T)$ when a vertex v gets added to T on account of condition (2a). The number of marked edges is clearly at least $bd|T_2|$ since the left end of each marked edge is in T_2 (by the definition of T_2) and each such left end contributes at least bd marked edges. On the other hand, by the definition of $\Gamma(T)$, each vertex in G is joined to at most k of its backward neighbours by a marked edge. Hence, $bd|T_2| \leq k|V|$. □

Consequently, since $(abd)^{1/2} < k \leq (abd)^{1/2} + 1$, we have

$$|T| \leq |T_1| + |T_2| \leq \frac{a|V|}{k} + \frac{k|V|}{bd}$$

$$\leq \frac{a|V|}{(abd)^{1/2}} + \frac{((abd)^{1/2}+1)|V|}{bd} \leq |V|\left(2\left(\frac{a}{bd}\right)^{1/2} + \frac{1}{bd}\right).$$

It remains to show that $\mu(\mathcal{C}) \leq 2\Delta(a/bd)^{1/2} + \Delta/bd + bd$. To see this, recall that $\mathcal{C}(T) = A \cup T$ and notice that

$$|E(G[\mathcal{C}(T)])| \leq \Delta|T| + |E(G[A])| \leq \Delta|T| + bd|V|.$$

To see the last inequality, i.e., $|E(G[A])| \leq bd|V|$, note that a vertex v is removed from A by our algorithm unless we have $|F(v) \setminus \Gamma(T)| < bd$ at the stage where we consider v. Since each member of $\Gamma(T)$ is (eventually) removed from A, we see that each vertex of A has at most bd forward neighbours in A and the inequality follows. The claimed bound for $\mu(\mathcal{C})$ then follows from our previously established upper bound for $|T|$. □

To prove Theorem 5.2, we now combine Lemma 5.3 with Proposition 5.1. First note that the Kneser graph $K(n,r)$ has $\mathbf{V} = n\mathbf{N}/r$ vertices and is $(n-r)\mathbf{M}/r$ regular.

Let us take $\hat{C}(\varepsilon) = 20/\varepsilon^2$. It is easy to check that given $\beta > 0$ and a family $\mathcal{A} \subset [n]^{(r)}$ with $e(\mathcal{A}) \leq m$, we can apply Lemma 5.3 with $a = m/\mathbf{V}$ and $b = \beta$ to get families $\mathcal{T} \subset [n]^{(r)}$ and $\mathcal{C}(\mathcal{T}) \subset [n]^{(r)}$ such that $\mathcal{T} \subset \mathcal{A} \subset \mathcal{C}(\mathcal{T})$,

$$\begin{aligned}|\mathcal{T}| &\leq 2\mathbf{V}\left(\frac{rm}{\beta(n-r)\mathbf{V}\mathbf{M}}\right)^{1/2} + \frac{r\mathbf{V}}{(n-r)\mathbf{M}} \\ &\leq 2\left(\frac{nm\mathbf{N}}{\beta(n-r)\mathbf{M}}\right)^{1/2} + \frac{n\mathbf{N}}{(n-r)\mathbf{M}} \leq k_1,\end{aligned}$$

and

$$\begin{aligned}e(\mathcal{C}(\mathcal{T})) \leq \mathbf{V}\mu(\mathcal{C}(\mathcal{T})) &\leq 2\left(\frac{(n-r)m\mathbf{V}\mathbf{M}}{r\beta}\right)^{1/2} + \frac{\mathbf{V}}{\beta} + \frac{(n-r)\beta\mathbf{V}\mathbf{M}}{r} \\ &\leq \mathbf{M}\left(2\left(\frac{n(n-r)m\mathbf{N}}{r^2\beta\mathbf{M}}\right)^{1/2} + \frac{n\mathbf{N}}{r\beta\mathbf{M}} + \frac{n(n-r)\beta\mathbf{N}}{r^2}\right) \leq k_2\mathbf{M}/2,\end{aligned}$$

where in the last inequality, we use the fact that $n(n-r)/r^2 \leq 1/2\varepsilon^2$.

Hence, by Proposition 5.1, we see that $|\mathcal{C}(\mathcal{T})| \le \mathbf{N} + k_2$. The theorem then follows by taking the families $\mathcal{C}(\mathcal{T})$ for every $\mathcal{T} \subset [n]^{(r)}$ with $|\mathcal{T}| \le k_1$. $\square$

5.2. Stability for the Kruskal–Katona theorem. An important ingredient in our proof of Theorem 1.1 is Lemma 3.4 which gives a uniform lower bound, using Theorem 2.2 and the Kruskal–Katona theorem, for $e(\mathcal{A})$ in terms of $\ell(\mathcal{A})$ when the size of $\mathcal{A}$ is large.

However, there is a price to be paid for proving such a uniform bound: the bound is quite poor for most families to which the lemma can be applied. Indeed, the families which are extremal for the argument in the proof of Lemma 3.4 must possess a great deal of structure. Instead of the Kruskal–Katona theorem, one should be able to use a stability version of the Kruskal–Katona theorem, as proved by Keevash [73] for example, to prove a more general result that accounts for the structure of the family under consideration.

6. Concluding remarks

Several problems related to the question considered here remain. First of all, it would be good to determine the largest possible value of $c(\varepsilon)$ with which Theorem 1.1 holds. It is likely that one needs new ideas to resolve this problem.

Second, one would also like to know what happens when r is very close to $n/2$. Perhaps most interesting is the case when $n = 2r + 1$; one would like to know the values of p for which we have $\alpha(K_p(2r+1, r)) = \binom{2r}{r-1}$ with high probability. A simple calculation shows that $p = 3/4$ is the threshold at which we are likely to find a star and an r-set not in the star all the edges between which are missing in $K_p(2r+1, r)$ which suggests that the critical threshold should be $3/4$. However, it would even be interesting to show that $\alpha(K_p(2r+1, r)) = \binom{2r}{r-1}$ with high probability for, say, all $p \ge 0.999$.

CHAPTER 7

Line percolation

Joint work with Paul Balister, Béla Bollobás and Jonathan Lee.

1. Introduction

Bootstrap percolation models and arguments have been used to study a range of phenomena in various areas, ranging from crack formation, clustering phenomena, the dynamics of glasses and sandpiles to neural nets and economics; see [84, 6, 53] for a small sample of such applications. In this chapter, we shall study a new geometric bootstrap percolation model defined on the d-dimensional grid $[n]^d$ with infection parameter $r \in \mathbb{N}$ which we call *r-neighbour line percolation*. Given $v \in [n]^d$, write $L(v)$ for the set of d axis parallel lines through v and let

$$L([n]^d) = \bigcup_{v \in [n]^d} L(v)$$

be the set of all axis parallel lines that pass through the lattice points of $[n]^d$. In line percolation, infection spreads from a subset $A \subset [n]^d$ of initially infected lattice points as follows: if there is a line $\mathcal{L} \in L([n]^d)$ with r or more infected lattice points on it, then every lattice point of $[n]^d$ on $\mathcal{L}$ gets infected. In other words, we have a sequence $A = A^{(0)} \subset A^{(1)} \subset \ldots A^{(t)} \subset \ldots$ of subsets of $[n]^d$ such that

$$A^{(t+1)} = A^{(t)} \cup \left\{ v \in [n]^d : \exists \mathcal{L} \in L(v) \text{ such that } |\mathcal{L} \cap A^{(t)}| \geq r \right\}.$$

The *closure* of A is the set $[A] = \bigcup_{t \geq 0} A^{(t)}$ of eventually infected points. We say that the process *terminates* when no more newly infected points are added, i.e., when $A^{(t)} = [A]$. If all the points of $[n]^d$ are infected when the process terminates, i.e., if $[A] = [n]^d$, then we say that A *percolates*.

The classical model of *r-neighbour bootstrap percolation on a graph* was introduced by Chalupa, Leath and Reich [37] in the context of disordered magnetic systems and has since been extensively studied not only by mathematicians but also by physicists and sociologists; for a small sample of papers, see, for instance, [1, 55, 64, 107]. In this model, a vertex of the graph gets infected if it has at least r previously infected neighbours in the graph. The model is usually studied in the random setting, where the main question is to determine the critical threshold at which percolation occurs. If the elements of the initially infected set are chosen independently at random, each with probability p, then one aims to determine the value p_c at which percolation becomes likely. In this regard, the r-neighbour bootstrap percolation model on $[n]^d$, with edges induced by the integer lattice $\mathbb{Z}^d$, has been the subject of large body of work; see [70, 15, 14], and the references therein.

On account of its inherent geometric structure, it is possible to construct other interesting bootstrap percolation models on the d-dimensional grid. In the past, this has involved endowing the grid with a graph structure other than the one induced by the integer lattice (a Cartesian product of paths).

In this direction, Holroyd, Liggett and Romik [71] considered r-neighbour bootstrap percolation on $[n]^2$ where the neighbourhood of a lattice point v is taken to be a 'cross' centred at v, consisting of $r-1$ points in each of the four axis directions. Sharp thresholds for a model with an anisotropic variant of these cross neighbourhoods were obtained recently by Duminil-Copin and van Enter [44]. Gravner, Hoffman, Pfeiffer and Sivakoff [65] studied the r-neighbour bootstrap percolation model on $[n]^d$ with the edges induced by the Hamming torus where $u, v \in [n]^d$ are adjacent if and only if $u - v$ has exactly one nonzero coordinate; the Hamming torus, in other words, is the Cartesian product of complete graphs, which is perhaps the second most natural graph structure on $[n]^d$ after the grid. They obtained bounds on the critical exponents (i.e.,

$\log_n(p_c)$) which are tight in the case $d = 2$ and for small values of the infection parameter when $d = 3$.

The line percolation model we consider is a natural variant of the bootstrap percolation model on the Hamming torus studied by Gravner, Hoffman, Pfeiffer and Sivakoff. However, we should note that while all the other models mentioned above are r-neighbour bootstrap percolation models on *some underlying graph*, the line percolation model *is not*.

Morally, line percolation is better thought of as coming from the very general *neighbourhood family percolation* model introduced by Bollobás, Smith and Uzzell [31]. In the neighbourhood family percolation model, one starts by specifying a homogeneous, finite collection of subsets of the grid for each point of the grid; a point of the grid becomes infected if all the points of some set in the collection associated with the point are previously infected. In their paper, Bollobás, Smith and Uzzell prove a classification theorem for two-dimensional neighbourhood family models and show that every such model is of one of three types: *supercritical*, *critical* or *subcritical*. In particular, they show that a model is supercritical if and only if there exist finite sets from which the infection can spread to the whole lattice. While line percolation cannot be described by associating a *finite* family of neighbourhoods with each point of the lattice, there do exist, as we shall see, finite sets from which the infection can spread to the whole lattice in the line percolation model, and our results about the critical probabilities of the line percolation model are in agreement with the general bounds for the critical probabilities of supercritical models proved in [31]. For some related work concerning subcritical models, see the paper of Balister, Bollobás, Przykucki and Smith [12].

2. Our results

In this chapter, our main aim is to investigate what happens in the line percolation model when the initial set $A = A_p \subset [n]^d$ of infected points is determined by randomly selecting points from $[n]^d$, each independently with probability p. It would be natural to determine the values of p for which percolation is likely to occur. Let $\vartheta_p(n,r,d)$ denote the probability that such a randomly chosen initial set A_p percolates. We note that $\vartheta_p(n,r,d)$ increases with p and define the *critical probability* $p_c(n,r,d)$ by setting

$$p_c(n,r,d) = \inf\{p : \vartheta_p(n,r,d) \geq 1/2\}.$$

The primary question of interest is to determine the asymptotic behaviour of $p_c(n,r,d)$ as $n \to \infty$ for every $d, r \in \mathbb{N}$. Note that when the infection parameter equals one, a set A of initially infected lattice points percolates if and only if $|A| > 0$ which implies that $p_c(n,1,d) = \Theta(n^{-d})$; we restrict our attention to $r \geq 2$.

Before we state our results, a few remarks about asymptotic notation are in order. We shall make use of standard asymptotic notation; the variable tending to infinity will always be n unless we explicitly specify otherwise. When convenient, we shall also make use of some notation (of Vinogradov) that might be considered non-standard: given functions $f(n)$ and $g(n)$, we write $f \ll g$ if $f = O(g)$, $f \gg g$ if $g = O(f)$, and $f \sim g$ if $f = (1+o(1))g$. Constants suppressed by the asymptotic notation are allowed to depend on the fixed infection parameter r, but of course, not on n or p.

In two dimensions, we are able to estimate the probability of percolation $\vartheta_p(n,r,2)$ up to constant factors for all $0 \leq p \leq 1$. We also determine $p_c(n,r,2)$ up to a factor of $1+o(1)$ as $n \to \infty$.

Theorem 2.1. *Fix $r, s \in \mathbb{N}$, with $r \geq 2$ and $0 \leq s \leq r-1$. Then as $n \to \infty$,*

$$\vartheta_p(n,r,2) = \Theta\Big(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\Big) \tag{11}$$

when $n^{-1-\frac{1}{r-s-1}} \ll p \ll n^{-1-\frac{1}{r-s}}$, with the convention that $n^{-1-\frac{1}{r-s-1}}$ is zero when $s = r-1$. Also, $\vartheta_p(n,r,2) = \Theta(1)$ when $p \gg n^{-1-\frac{1}{r}}$. Furthermore,

$$p_c(n,r,2) \sim \lambda n^{-1-\frac{1}{r}}$$

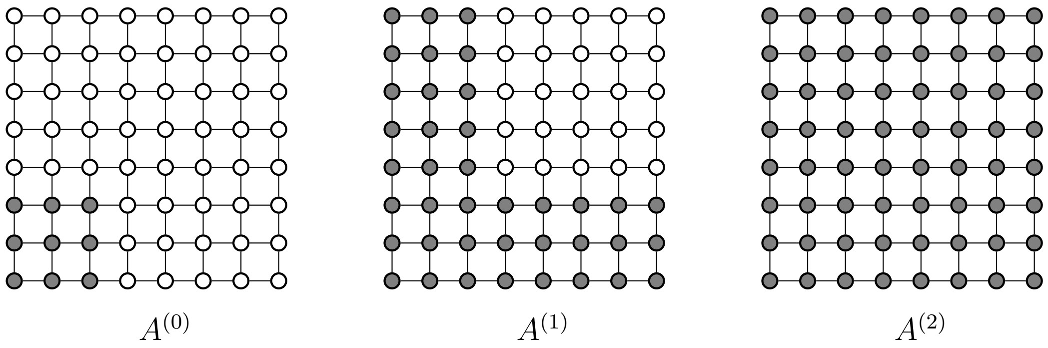

FIGURE 1. The spread of infection from $A = [3]^2$ in the 3-neighbour line percolation process on $[8]^2$.

where λ is the unique positive real number satisfying $\exp(-2\lambda^r/r!) = 1/2$.

The techniques used to obtain the above formula for $\vartheta_p(n, r, 2)$ allow us to prove the following result about the critical probability in three dimensions, which is the main result of this chapter.

THEOREM 2.2. *Fix $r \in \mathbb{N}$, with $r \geq 2$, and let $s = \lfloor \sqrt{r + 1/4} - 1/2 \rfloor$. Then as $n \to \infty$,*

$$p_c(n, r, 3) = \Theta\Big(n^{-1-\frac{1}{r-\gamma}}\Big)$$

where $\gamma = (r + s^2 + s)/2(s + 1)$.

The nature of the threshold at the critical probability is also worth investigating. We say that the model exhibits a sharp threshold at $p_c = p_c(n, r, d)$ if for any fixed $\varepsilon > 0$, $\vartheta_{(1+\varepsilon)p_c}(n, r, d) = 1 - o(1)$ and $\vartheta_{(1-\varepsilon)p_c}(n, r, d) = o(1)$. It is not difficult to see from our proofs of Theorems 2.1 and 2.2 that in stark contrast to the classical r-neighbour bootstrap percolation model on the grid, there is *no sharp threshold* at p_c when $d = 2, 3$. We expect similar behaviour in higher dimensions but we do not have a proof of such an assertion.

It is also an interesting question to determine the size of a minimal percolating set for r-neighbour line percolation on $[n]^d$ for any $d, r \in \mathbb{N}$ and $n \geq r$. It is easy to check that the set $[r]^d$ percolates (see Figure 1). We shall demonstrate that this is in fact optimal.

THEOREM 2.3. *Let $d, r, n \in \mathbb{N}$, with $n \geq r$. Then the minimum size of a percolating set in the r-neighbour line percolation process on $[n]^d$ is r^d.*

Establishing this fact turns out to be harder than it appears at first glance. The result is trivial when $d = 1$. When $d = 2$, it is not hard to demonstrate that any percolating set has size at least r^2. Consider a generalised two-dimensional line percolation model on $[n]^2$ where the infection thresholds for horizontal and vertical lines are r_h and r_v respectively; we recover the r-neighbour line percolation model when $r_h = r_v = r$. Let $M(r_h, r_v)$ denote the size of a minimal percolating set in this generalised model. Consider the first line $\mathcal{L}$ to be infected: if $\mathcal{L}$ is horizontal, then $\mathcal{L}$ must contain r_h initially infected points and furthermore, if the set of initially infected points is a percolating set, then the set of initially infected points not on $\mathcal{L}$ must constitute a percolating set for the generalised process with infection parameters r_h and $r_v - 1$. An analogous statement holds if $\mathcal{L}$ is vertical. It follows that

$$M(r_h, r_v) \geq \min(r_v + M(r_h - 1, r_v), r_h + M(r_h, r_v - 1)).$$

We obtain by induction that $M(r_h, r_v) \geq r_h r_v$ which implies in particular that $M(r, r) \geq r^2$. The argument described above depends crucially on the fact that a line has codimension one in a two-dimensional space. The incidence geometry of a collection of lines in the plane is essentially straightforward; this is no longer the case in higher dimensions and we need more delicate arguments to prove Theorem 2.3.

The rest of this chapter is organised as follows. We collect together some useful facts about binomial random variables in Section 3. We consider line percolation in two dimensions in Section 4, and prove Theorem 2.1. In Section 5, we turn to line percolation in three dimensions and prove Theorem 2.2, thus obtaining an estimate for the critical probability which is tight up to multiplicative constants. In Section 6, we determine the size of minimal percolating sets for r-neighbour line percolation on $[n]^d$ and prove Theorem 2.3. We conclude the chapter in Section 7 with some discussion.

3. Probabilistic preliminaries

We will need some standard facts about binomial random variables. We collect these here for the sake of convenience. As is usual, for a random variable with distribution $\mathrm{Bin}(N, p)$, we write $\mu = Np$ for its mean.

The first proposition we shall require is an easy consequence of the fact that $e^{-2x} \le 1 - x \le e^{-x}$ for all $0 \le x \le 1/2$.

PROPOSITION 3.1. *Let X be a random variable with distribution* $\mathrm{Bin}(N, p)$, *with $p \le 1/2$. Then for any $1 \le k \le n$,*

$$\exp(-2\mu)(\mu/k)^k \le \mathbb{P}(X = k) \le \exp(-\mu)(2e\mu/k)^k.$$

Also, $\exp(-2\mu) \le \mathbb{P}(X = 0) \le \exp(-\mu)$.

PROOF. We have $\mathbb{P}(X = k) = \binom{N}{k} p^k (1-p)^{N-k}$. The required bounds follow from the fact that $\exp(-2p) \le 1 - p \le \exp(-p)$ for $0 \le p \le 1/2$ and the fact that $(N/k)^k \le \binom{N}{k} \le (eN/k)$ for all $k \ge 1$. $\square$

The following proposition is an immediate corollary of Proposition 3.1.

PROPOSITION 3.2. *Let X be a random variable with distribution* $\mathrm{Bin}(N, p)$. *Then for any fixed $k \ge 0$, as $N \to \infty$,*

$$\mathbb{P}(X \ge k) = \begin{cases} \Theta(\mathbb{P}(X = k)) = \Theta(\mu^k) & \text{if } \mu \ll 1, \\ \Theta(1) & \text{if } \mu \gg 1. \end{cases}$$

where $\mu = Np$ is the mean of X.

We shall also make use of the following standard concentration result; see Appendix A of [5] for example.

PROPOSITION 3.3. *Let X be a random variable with distribution* $\mathrm{Bin}(N, p)$. *Then for any $0 < \delta < 1$,*

$$\mathbb{P}(|X - \mu| > \delta\mu) \le \exp\left(\frac{-\delta^2 \mu}{3}\right). \qquad \square$$

Finally, we shall also need the following simple consequence of Harris's Lemma. Given a set $A \subset [n]$, we say that $E \subset \{0,1\}^{[n]}$ is *decreasing on* A if $\omega \in E$ implies that $\omega' \in E$ for all $\omega' \in \{0,1\}^{[n]}$ such that $\omega'_x \le \omega_x$ for $x \in A$ and $\omega'_x = \omega_x$ for all $x \notin A$.

PROPOSITION 3.4. *Let $A \subset [n]$ and let $\mathbb{P}$ be a product measure on $\{0,1\}^{[n]}$. Then for any increasing event F which depends only on the coordinates in A and any event E which is decreasing on A, we have $\mathbb{P}(F|E) \le \mathbb{P}(F)$.*

PROOF. For $v \in \{0,1\}^{[n]\setminus A}$, denote by I_v the event that the coordinates in $[n] \setminus A$ are given by v. Then, since E is decreasing on A and F is increasing on A, we see by applying Harris's Lemma to the induced product measure on $\{0,1\}^A$ that $\mathbb{P}(E \cap F \mid I_v) \le \mathbb{P}(E \mid I_v)\mathbb{P}(F \mid I_v)$ for every $v \in \{0,1\}^{[n]\setminus A}$. Since F does not depend on the coordinates in $[n] \setminus A$, we also have $\mathbb{P}(F \mid I_v) = \mathbb{P}(F)$ for every $v \in \{0,1\}^{[n]\setminus A}$. Therefore, by summing over all $v \in \{0,1\}^{[n]\setminus A}$, we see that

$$\begin{aligned}\mathbb{P}(E \cap F) &= \sum_v \mathbb{P}(I_v)\mathbb{P}(E \cap F \mid I_v)\\ &\le \sum_v \mathbb{P}(I_v)\mathbb{P}(E \mid I_v)\mathbb{P}(F \mid I_v)\\ &= \sum_v \mathbb{P}(I_v)\mathbb{P}(E \mid I_v)\mathbb{P}(F) = \mathbb{P}(E)\mathbb{P}(F)\end{aligned}$$

and the proposition follows. □

4. Line percolation in two dimensions

The proof of the following proposition is essentially identical to the proof of Theorem 2.1 in [65]; we reproduce it here for completeness.

PROPOSITION 4.1. *Fix $r \in \mathbb{N}$, with $r \ge 2$, and let $C > 0$ be a positive constant. If $p \sim Cn^{-1-1/r}$, then*

$$\vartheta_p(n, r, 2) \sim 1 - \exp\left(-2C^r/r!\right).$$

PROOF. The probability that a given line has $r+1$ or more initially infected points on it is bounded above by $\binom{n}{r+1}p^{r+1}$ which implies that the probability that any line has $r+1$ or more initially infected points on it is bounded above by

$$2n\binom{n}{r+1}p^{r+1} = O\big(n^{r+2}p^{r+1}\big) = O\big(n^{-1/r}\big).$$

Consequently, with high probability, no line has $r+1$ or more initially infected points on it.

Let E_h denote the event that some horizontal line contains r initially infected points and define E_v analogously. Clearly, the process terminates on the first step if neither E_h nor E_v hold; so $\vartheta_p \leq \mathbb{P}(E_h \cup E_v)$.

The number of horizontal lines with r initially infected points is binomially distributed (since the events corresponding to distinct horizontal lines are independent) and it is easily seen to converge in distribution to a Poisson random variable with mean $C^r/r!$ since the expected number of such horizontal lines is $n\binom{n}{r}p^r(1-p)^{n-r} \sim C^r/r!$. Thus $\mathbb{P}(E_h) \sim 1-\exp(-C^r/r!)$; similarly, $\mathbb{P}(E_v) \sim 1-\exp(-C^r/r!)$.

We now estimate $\mathbb{P}(E_h \cap E_v)$. Let $E_h \circ E_v$ denote the event that E_h and E_v occur disjointly. Now, E_h and E_v are increasing events, and so it follows from Harris's Lemma and the BK inequality that $\mathbb{P}(E_h \cap E_v) \geq \mathbb{P}(E_h)\mathbb{P}(E_v) \geq \mathbb{P}(E_h \circ E_v)$. Observe that $(E_h \cap E_v) \setminus (E_h \circ E_v)$ happens only if some lattice point v is initially infected and each of the two axis parallel lines through v contain $r-1$ initially infected points. It follows that

$$\mathbb{P}((E_h \cap E_v) \setminus (E_h \circ E_v)) = O\big(n^2p(np)^{2r-2}\big) = O\big(n^{-1+1/r}\big)$$

and so $\mathbb{P}((E_h \cap E_v) \setminus (E_h \circ E_v)) = o(1)$. Consequently, we see that $\mathbb{P}(E_h \cap E_v) \sim \mathbb{P}(E_h)\mathbb{P}(E_v)$. Hence,

$$\mathbb{P}(E_h \cup E_v) \sim \mathbb{P}(E_h) + \mathbb{P}(E_v) - \mathbb{P}(E_h)\mathbb{P}(E_v) \sim 1 - \exp\left(-2C^r/r!\right)$$

and so $\vartheta_p \leq 1 - \exp\left(-2C^r/r!\right) + o(1)$.

To bound ϑ_p from below, we generate the initial configuration in two rounds, first by sprinkling infected points with density $p' = p(1 - 1/\log n)$ and then (independently) with density $p'' = p/\log n$; clearly, this configuration is dominated by an initial configuration where points are infected independently with density p, and so it suffices to lower bound the probability of percolation starting from this configuration.

Let E' be the event that some line contains r initially infected points from the *first sprinkling* of points. It is easy to check from the argument above that $\mathbb{P}(E') \sim 1 - \exp(-2C^r/r!)$.

Let us now condition on E' and suppose that some line $\mathcal{L}$ has r infected points on it from the first sprinkling. The probability that a particular line perpendicular to $\mathcal{L}$ has $r-1$ initially infected points from the *second sprinkling* of points (none of which are on $\mathcal{L}$) is $\Theta(((n-1)p'')^{r-1}) = \Theta(n^{-1+1/r}(\log n)^{-r+1})$. Thus, the number of such lines is a binomial random variable with mean $\mu = \Omega(n^{1/r}(\log n)^{-r})$. Since $\mu \to \infty$ as $n \to \infty$, by Proposition 3.3, the probability that there exist at least r such lines in the second sprinkling is $1 - o(1)$. Hence conditional on E', the probability of percolating using the points infected in the second round is $1 - o(1)$. Hence

$$\vartheta_p \geq (1 - o(1))\mathbb{P}(E') = 1 - \exp\left(-2C^r/r!\right) - o(1)$$

and the result follows. $\square$

We shall now prove Theorem 2.1.

Proof of Theorem 2.1. It follows from Proposition 4.1 that $p_c(n, r, 2) \sim \lambda n^{-1-1/r}$ where λ is the unique positive real number satisfying $\exp(-2\lambda^r/r!) = 1/2$. So we know that $\vartheta_p(n, r, 2) = \Theta(1)$ when $p \gg n^{-1-1/r}$.

Let us now suppose that $p \ll n^{-1-1/r}$. We fix $s \in \{0, 1, \ldots, r-1\}$ to be the least natural number for which $n(np)^{r-(s+1)} \gg 1$; hence $n(np)^{r-i} \ll 1$ for each $0 \leq i \leq s$.

We first bound $\vartheta_p(n,r,2)$ from below. To do so, we set $p' = p/(2s+2)$ and generate our initial configuration by sprinkling infected points in $2s+2$ rounds where we infect points independently with probability p' in each of these rounds; such a configuration is clearly dominated by a configuration where we infect each point independently with probability p, so a lower bound for the probability of percolating from such a configuration is also a lower bound for ϑ_p.

Let us estimate the probability that we can find distinct lines $\mathcal{L}_1, \ldots, \mathcal{L}_{2s+1}$ such that

(1) For $0 \le j \le s$, the line $\mathcal{L}_{2j+1}$ is a horizontal line containing $r-j$ infected points from the sprinkling in round $2j+1$, none of which lie on $\mathcal{L}_2, \mathcal{L}_4, \ldots, \mathcal{L}_{2j}$, and
(2) for $1 \le j \le s$, the line $\mathcal{L}_{2j}$ is a vertical line containing $r-j$ infected points from the sprinkling in round $2j$, none of which lie on $\mathcal{L}_1, \mathcal{L}_3, \ldots, \mathcal{L}_{2j-1}$.

Indeed, conditional on finding $\mathcal{L}_1, \ldots, \mathcal{L}_{2j}$ in the first $2j$ rounds, the probability that a given horizontal line has $r-j$ infected points on it from the sprinkling in round $2j+1$, none of which lie on $\mathcal{L}_2, \mathcal{L}_4, \ldots, \mathcal{L}_{2j}$ is clearly $\Theta(((n-j)p')^{r-j}) = \Theta((np)^{r-j})$. Since $j \le s$, we have $n(np)^{r-j} \ll 1$, so the probability that we can find a suitable horizontal line $\mathcal{L}_{2j+1}$ distinct from $\mathcal{L}_1, \ldots, \mathcal{L}_{2j-1}$ in round $2j+1$ is $\Theta((n-j+1)(np)^{r-j}) = \Theta(n(np)^{r-j})$ by Proposition 3.2. The probability that we can find $\mathcal{L}_{2j}$ as required conditional on finding $\mathcal{L}_1, \ldots, \mathcal{L}_{2j-1}$ is similarly seen to be $\Theta(n(np)^{r-j})$.

So the probability that we can find $\mathcal{L}_1, \ldots, \mathcal{L}_{2s+1}$ as required in the first $2s+1$ rounds is thus, up to constant factors, at least

$$n(np)^r \times n(np)^{r-1} \times \cdots \times n(np)^{r-s} \times n(np)^{r-s} = n^{2s+1}(np)^{r(2s+1)-s(s+1)}.$$

Conditional on finding $\mathcal{L}_1, \ldots, \mathcal{L}_{2s+1}$, we claim that with probability $\Theta(1)$, there are at least r vertical lines distinct from $\mathcal{L}_2, \ldots, \mathcal{L}_{2s}$ each containing

$r - s - 1$ infected points from the sprinkling in round $2s + 2$, none of which lie on $\mathcal{L}_1, \mathcal{L}_3, \ldots, \mathcal{L}_{2s+1}$. Indeed, the number of such lines is binomially distributed with mean $\mu = \Omega(n(np)^{r-s-1})$, and we know that $\mu \gg 1$ by the definition of s, so we are done by Proposition 3.2.

Such a configuration clearly percolates. Indeed, first the line $\mathcal{L}_1$ gets infected, then the line $\mathcal{L}_2$, and so on, until all of $\mathcal{L}_1, \ldots, \mathcal{L}_{2s+1}$ are infected. At this point, the r vertical lines distinct from $\mathcal{L}_2, \ldots, \mathcal{L}_{2s}$ each containing $r - s - 1$ infected points none of which lie on $\mathcal{L}_1, \mathcal{L}_3, \ldots, \mathcal{L}_{2s+1}$ get infected. Of course, once we have r parallel fully infected lines, every point gets infected. Thus, we deduce from this sprinkling argument that $\vartheta_p = \Omega(n^{2s+1}(np)^{r(2s+1)-s(s+1)})$.

We now give a matching upper bound for $\vartheta_p(n, r, 2)$. To do so, it will be convenient to work with a modified two-dimensional line percolation process

$$A_p = G^{(0)} \subset G^{(1)} \subset \ldots G^{(t)} \subset \ldots$$

where $G^{(2t+1)}$ is obtained from $G^{(2t)}$ by spreading the infection (only) along horizontal lines and $G^{(2t+2)}$ is obtained from $G^{(2t+1)}$ by spreading the infection along vertical lines. Since $G^{(t)} \subset A^{(t)}$ and $A^{(t)} \subset G^{(2t)}$, percolation occurs in the original process if and only if it occurs in the modified process.

Note that A_p percolates if and only if some $G^{(t)}$ contains r or more fully infected lines; indeed, in this case $G^{(t+2)} = [n]^2$. In particular, since $s + 1 \leq r$, if A_p percolates, then some $G^{(t)}$ contains $s + 1$ parallel fully infected lines. Let us stop the modified process as soon it produces $s + 1$ or more parallel fully infected lines (or reaches termination). Note that we necessarily stop the process in at most $2s + 2 \leq 2r$ steps.

Let us clarify what we mean by a 'fully infected line'. Formally, a line becomes fully infected when we examine its direction and find that there are r or more infected points on the line, and hence infect the rest of the line. Note that, unlikely as it might be, all the points on a line could become infected before we inspect the direction corresponding to the line; in this case, we declare the line to be fully infected only *after* we examine its direction.

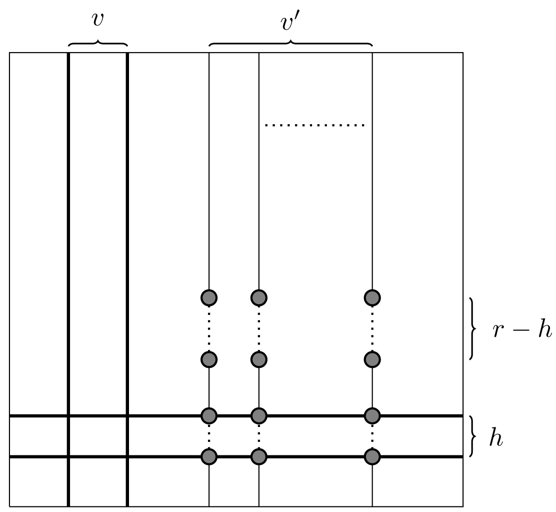

FIGURE 2. The conditioning argument.

Let E^* denote the event that we stop the process because it generated $s+1$ parallel fully infected lines; clearly $\vartheta_p \le \mathbb{P}(E^*)$.

The following lemma will allow us to bound $\mathbb{P}(E^*)$ from above.

LEMMA 4.2. *Let $t, h, v \in \mathbb{N}$ and suppose that t is odd and that $(n(np)^{r-h}) \ll 1$. Let $\mathcal{E}$ be the event that the numbers of fully infected horizontal and vertical lines after the first t steps are h and v respectively and $\mathcal{F}$ denote the event that the process generates v' or more vertical lines on step $t+1$. Then*

$$\mathbb{P}(\mathcal{F} \mid \mathcal{E}) = O\Big(\Big(n(np)^{r-h}\Big)^{v'}\Big).$$

PROOF. We prove this with a conditioning argument. Consider any set $\mathcal{H}$ of h horizontal lines and any set $\mathcal{V}$ of v vertical lines and fix a partition $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_3 \cup \cdots \cup \mathcal{H}_t$ and $\mathcal{V} = \mathcal{V}_2 \cup \mathcal{V}_4 \cup \cdots \cup \mathcal{V}_{t-1}$. Let E be the event that the lines infected in the first t steps of the process are precisely those in $\mathcal{H}_1, \mathcal{V}_2, \ldots, \mathcal{H}_t$. Since $\mathcal{E}$ is the disjoint union of such events E, it suffices to show that $\mathbb{P}(\mathcal{F} \mid E) = O((n(np)^{r-h})^{v'})$ for every such event E.

Let $\mathcal{S}$ denote the subgrid consisting of those points not on any of the lines in $\mathcal{H} \cup \mathcal{V}$. Let F denote the event that there are v' or more vertical lines in $\mathcal{S}$ with $r-h$ or more initially infected points on them. Note that conditional on E, the event $\mathcal{F}$ occurs if and only if F occurs; clearly, $\mathbb{P}(\mathcal{F} \mid E) = \mathbb{P}(F \mid E)$. Note that F is an increasing event and $\mathbb{P}(F) = O((n(np)^{r-h})^{v'})$ since $(n(np)^{r-h}) \ll 1$.

If E were a decreasing event, we could conclude immediately from the Harris's Lemma that $\mathbb{P}(F \mid E) \leq \mathbb{P}(F)$. Though E is in itself not decreasing, we claim that E is decreasing on $\mathcal{S}$.

Let $\omega \in \{0,1\}^{[n]^2}$ be an initial configuration that belongs to E and let $\omega' \in \{0,1\}^{[n]^2}$ be such that $\omega'_x \leq \omega_x$ for $x \in \mathcal{S}$ and $\omega'_x = \omega_x$ for $x \notin \mathcal{S}$. We now check that $\omega' \in E$. We first note that since the percolation process is monotone and $\omega' \leq \omega$, the set of fully infected lines at any stage when we start from ω' is a subset of the set of fully infected lines at that stage when we start from ω. Next, note that since $\omega \in E$, when we start the line percolation process from ω, none of the lines through any of the points in $\mathcal{S}$ are infected after the first t steps; in other words, none of these points participate in the spread of infection during the first t steps. Therefore, since $\omega'_x = \omega_x$ for all $x \notin \mathcal{S}$, it follows that the set of fully infected lines after the first t steps when we start from ω' is actually identical to the set of fully infected lines after the first t steps when we start from ω. Thus, $\omega' \in E$ and it follows that E is decreasing on $\mathcal{S}$.

It follows by Proposition 3.4 that $\mathbb{P}(F \mid E) \leq \mathbb{P}(F)$; the lemma follows. $\square$

Let h_t and v_t be the numbers of horizontal and vertical lines infected when going from $G^{(2t)}$ and $G^{(2t+1)}$ and from $G^{(2t+1)}$ to $G^{(2t+2)}$ respectively; the pair $(\mathbf{h} = (h_t)_{t\geq 0}, \mathbf{v} = (v_t)_{t\geq 0})$ is called the *line-count* of the modified percolation process.

Given two sequences $\mathbf{h} = (h_t)_{t=0}^{k}$ and $\mathbf{v} = (v_t)_{t=0}^{k}$ of positive integers with $k \leq s+1$, we say that $(\mathbf{h}, \mathbf{v})$ is a *vertical line-count* if $(\mathbf{h}, \mathbf{v})$ is the line-count of a process which generates $s+1$ fully infected vertical lines before it generates $s+1$ fully infected horizontal lines, and does so in exactly $2k+2$ steps; in other words, if

(1) $\sum_{t<k} v_t \leq s$,
(2) $\sum_{t\leq k} h_t \leq s$, and

(3) $\sum_{t\le k} v_t \ge s+1$.

The definition of a *horizontal line-count* $(\mathbf{h} = (h_t)_{t=0}^{k+1}, \mathbf{v} = (v_t)_{t=0}^{k})$ is analogous. Given a vertical line-count $(\mathbf{h} = (h_t)_{t=0}^{k}, \mathbf{v} = (v_t)_{t=0}^{k})$, let us define its *(vertical) preface* to be the pair $(\mathbf{h}, \mathbf{v}')$ where $\mathbf{v}' = (v_t)_{t=0}^{k-1}$. Similarly, the *(horizontal) preface* of a horizontal line-count $(\mathbf{h} = (h_t)_{t=0}^{k+1}, \mathbf{v} = (v_t)_{t=0}^{k})$ is the pair $(\mathbf{h}', \mathbf{v})$ where $\mathbf{h}' = (h_t)_{t=0}^{k}$.

Given a vertical preface $(\mathbf{h}, \mathbf{v}')$, let $E_v(\mathbf{h}, \mathbf{v}')$ be the event that the process generates $s+1$ fully infected vertical lines before it generates $s+1$ fully infected horizontal lines and furthermore, the (vertical) line-count of the process has preface $(\mathbf{h}, \mathbf{v}')$. For a horizontal preface $(\mathbf{h}', \mathbf{v})$, define $E_h(\mathbf{h}', \mathbf{v})$ analogously. We then note that

$$\mathbb{P}(E^*) = \sum_{(\mathbf{h},\mathbf{v}')} \mathbb{P}(E_v(\mathbf{h}, \mathbf{v}')) + \sum_{(\mathbf{h}',\mathbf{v})} \mathbb{P}(E_h(\mathbf{h}', \mathbf{v}))$$

where the two sums are over all valid vertical and horizontal prefaces respectively.

To specify a valid preface, we need to specify at most $2(s+1)$ distinct positive integers, each of which is at most $s+1$. So the number of valid prefaces is at most r^{2r}; consequently, to estimate $\mathbb{P}(E^*)$ up to constant factors, it suffices to estimate the largest of the probabilities $\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}'))$ and $\mathbb{P}(E_h(\mathbf{h}', \mathbf{v}))$.

We see by repeatedly applying Lemma 4.2 that $\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}'))$, up to constant factors, is bounded above by

$$(n(np)^r)^{h_0} \times (n(np)^{r-h_0})^{v_0} \times (n(np)^{r-v_0})^{h_1} \times \dots$$
$$\dots \times (n(np)^{r-\sum_{t<k} v_t})^{h_k} \times (n(np)^{r-\sum_{t\le k} h_t})^{s+1-\sum_{t<k} v_t}$$

which, on algebraic simplification, is seen to be $n^{s+1+h}(np)^{r(s+1)+(r-s-1)h}$ where $h = \sum_{t\le k} h_t$. To see this, set $v_k = s+1-\sum_{t<k} v_t$ and note that the exponent of n in the expression above (not including those factors of n coming from the

powers of np) is

$$\sum_{t\le k} h_t + \sum_{t\le k} v_t = h + s + 1$$

and the exponent of np is

$$\begin{aligned} r\left(\sum_{t\le k} h_t + \sum_{t\le k} v_t\right) - \sum_{t',t''\le k} h_{t'}v_{t''} &= r(s+1+h) - \left(\sum_{t\le k} h_t\right)\left(\sum_{t\le k} v_t\right) \\ &= r(s+1+h) - (s+1)h \\ &= r(s+1) + (r-s-1)h \end{aligned}$$

as required. Since $n(np)^{r-s-1} \gg 1$, this upper bound increases with h, and since $h \le s$, it is maximised when $h = s$. Thus, we see that

$$\mathbb{P}(E_v(\mathbf{h}, \mathbf{v}')) = O\Big(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\Big)$$

for any valid vertical preface. The same bound applies for each horizontal preface, and it follows that $\vartheta_p \le \mathbb{P}(E^*) = O(n^{2s+1}(np)^{r(2s+1)-s(s+1)})$. $\square$

5. The critical probability in three dimensions

We now turn our attention to the line percolation process in three dimensions. We shall now prove Theorem 2.2.

Proof of Theorem 2.2. We prove the upper and lower bounds separately. Suppose that points are initially infected independently with probability p and set $C = C(n) = p/n^{-1-1/(r-\gamma)}$. We distinguish two cases.

Case 1: $C \gg 1$. Unsurprisingly, it is easier to show that percolation occurs than to demonstrate otherwise. We start by bounding p_c from above by showing that percolation occurs with probability at least $1/2$ if C is greater than some sufficiently large constant. Note that $s = \lfloor\sqrt{r+1/4} - 1/2\rfloor$, as defined in the statement of Theorem 2.2, is the greatest natural number such that $s(s+1) \le r$. Since $s(s+1) \le r$ and $(s+1)(s+2) > r$, it is not hard to check that $\gamma = (r + s^2 + s)/(2s + 2)$ satisfies

$$r - s - 1 \leq \gamma \leq r - s$$

and hence

$$n^{-1-\frac{1}{r-s-1}} \ll n^{-1-\frac{1}{r-\gamma}} \ll n^{-1-\frac{1}{r-s}}, \tag{12}$$

and so it follows from Theorem 2.1 that

$$\vartheta_p(n, r, 2) = \Theta\Big(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\Big) = \Theta\big(C^{r(2s+1)-s(s+1)}n^{-1}\big).$$

We say that a plane $\mathcal{P}$ is *internally spanned* if $A^{(0)} \cap \mathcal{P}$ percolates in the line percolation process restricted to $\mathcal{P}$. Choose any direction and consider the n (parallel) planes perpendicular to this direction. The number of such planes which are internally spanned is a binomial random variable with mean $\mu = \Omega(C^{r(2s+1)-s(s+1)})$. Since $\mu \to \infty$ as $C \to \infty$, we see from Proposition 3.3 that there exist r parallel internally spanned planes with probability at least $1/2$ if C is greater than some sufficiently large constant. So it follows that $p_c(n, r, 3) = O(n^{-1-1/(r-\gamma)})$.

Case 2: $C \ll 1$**.** We claim that the probability of percolation is at most $1/2$, provided C is less than some sufficiently small constant.

We shall demonstrate that the probability of percolation is at most $1/2$ by proving something much stronger. We shall track, as the infection spreads, the number of planes with k or more parallel fully infected lines for each $1 \leq k \leq s + 1$ and show that, with probability at least $1/2$, these numbers are not too large when the process terminates; in particular, we shall show that there are no planes with $s + 1$ or more parallel fully infected lines when the process reaches termination and consequently, that there is no percolation.

As we did in the two-dimensional case, we shall work with a modified three-dimensional line percolation process in which the infection spreads along a single direction at each step. More precisely, denoting the standard basis for

$\mathbb{R}^3$ by $\{e_0, e_1, e_2\}$, in the modified process, we have a sequence

$$A_p = H^{(0)} \subset H^{(1)} \subset \cdots \subset H^{(t)} \subset \ldots$$

of subsets of $[n]^3$ where $H^{(t+1)}$ is obtained from $H^{(t)}$ by spreading the infection (only) along lines parallel to e_i where $i \equiv t \pmod 3$. Clearly, $H^{(t)} \subset A^{(t)} \subset H^{(3t)}$ and so A_p percolates in the original process if and only if it percolates in this modified process.

We run the modified three-dimensional process starting from A_p and we stop the process after step t if either

(A) the number of planes containing k or more parallel fully infected lines now exceeds $n^{1-k\gamma/(r-\gamma)}$ for some $1 \leq k \leq s+1$, or
(B) the process has terminated, i.e., $H^{(t)} = [A_p]$.

Let E_A be the event that we the stop the modified process on account of condition (A).

LEMMA 5.1. *In the modified process,*

$$\mathbb{P}(E_A) = O\left(\sum_{1 \leq k \leq s} C^{rk} + C^{r(2s+1)-s(s+1)}\right).$$

Let us fix a plane $\mathcal{P}$. For concreteness, we shall assume that $\mathcal{P}$ contains lines parallel to e_0 and e_1, and that e_2 is perpendicular to $\mathcal{P}$. We shall prove Lemma 5.1 by estimating the probability that $\mathcal{P}$ contains k or more parallel fully infected lines when we stop the process.

The spread of infection within $\mathcal{P}$ resembles the two-dimensional line percolation process, with the key distinction that some points in $\mathcal{P}$ also become infected *during* the process by virtue of lying on a fully infected line perpendicular to $\mathcal{P}$. However, since we are interested in estimating the probability that k or more parallel lines in $\mathcal{P}$ get infected *before* we stop the process, we shall not have to worry about there being too many such points.

For $1 \leq k \leq s+1$, let E_k denote the event that k or more parallel lines in $\mathcal{P}$ get infected before we stop the process. We shall prove Lemma 5.1 by bounding $\mathbb{P}(E_k)$ from above.

As before, we shall estimate the probability that k or more parallel lines in $\mathcal{P}$ get infected before we stop the process by estimating the probability that this happens according to a particular line-count. Note that unlike in the two-dimensional process, a large number of steps may elapse before a new line in $\mathcal{P}$ gets infected. Consequently, the precise notion of a line-count that we use here differs slightly from the notion used previously.

We call a step of the modified three-dimensional process an *epoch* if a non-empty subset of the lines in $\mathcal{P}$ (along a particular direction) get infected on that step. A *line-count* $\boldsymbol{\ell} = ((l_i, d_i))_{i \geq 0}$ is a sequence of pairs (l_i, d_i) such that each l_i is a positive integer and each $d_i \in \{e_0, e_1\}$ is a direction, with the property that either

(1) $\sum_{i:d_i=e_0} l_i \geq k$, or
(2) $\sum_{i:d_i=e_1} l_i \geq k$.

Given a line-count $\boldsymbol{\ell} = ((l_i, d_i))_{i \geq 0}$, we define its preface $\boldsymbol{\ell}^* = ((l_i, d_i))_{i=0}^{m}$ by taking m to be largest index j such that

(1) $\sum_{i \leq j:d_i=e_0} l_i < k$, and
(2) $\sum_{i \leq j:d_i=e_1} l_i < k$.

Note that if E_k occurs, there is a line-count $\boldsymbol{\ell} = ((l_i, d_i))_{i \geq 0}$ such that the number and direction of the lines infected on the i-th epoch are given by l_i and d_i respectively. Consequently, given a preface $\boldsymbol{\ell}^*$, let us write $E_k(\boldsymbol{\ell}^*)$ for the event that E_k occurs and that the preface of the associated line-count is $\boldsymbol{\ell}^*$. Clearly,

$$\mathbb{P}(E_k) = \sum_{\boldsymbol{\ell}^*} \mathbb{P}(E_k(\boldsymbol{\ell}^*)),$$

where the sum above is over all valid prefaces. Since the length of a valid preface is at most $2k$, the number of valid prefaces is at most $(2k)^{2k}$. Therefore, it suffices to bound the largest of the probabilities $\mathbb{P}(E_k(\boldsymbol{\ell}^*))$.

The following lemma allows us to estimate $\mathbb{P}(E_k(\boldsymbol{\ell}^*))$. We think of the lines in $\mathcal{P}$ parallel to e_0 as being horizontal, and the ones parallel to e_1 as being vertical.

LEMMA 5.2. *Let $i, h, v \in \mathbb{N}$ and suppose that $n(np)^{r-h} \ll 1$. Let $\mathcal{E}$ be the event that after the i-th epoch, the numbers of fully infected horizontal and vertical lines in $\mathcal{P}$ are h and v respectively. Also, let $\mathcal{F}$ denote the event that v' or more vertical lines in $\mathcal{P}$ get infected on the $(i+1)$-th epoch. Then*

$$\mathbb{P}(\mathcal{F} \,|\, \mathcal{E}) = O\bigg(\Big(n(np)^{r-h}\Big)^{v'}\bigg).$$

PROOF. The proof is very similar to that of Lemma 4.2. However, we also need to account for points that might not be infected initially but which get infected at some stage before the $(i+1)$-th epoch. We call a point of $\mathcal{P}$ a *boost* if the line perpendicular to $\mathcal{P}$ through that point is fully infected before the $(i+1)$-th epoch.

Let $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_i$ be disjoint non-empty sets of parallel lines in $\mathcal{P}$ such that the numbers of horizontal and vertical lines in $\mathcal{L}_1 \cup \mathcal{L}_2 \cup \cdots \cup \mathcal{L}_i$ are h and v respectively. Let $\mathcal{S}$ denote the subgrid consisting of those points of $\mathcal{P}$ not on any of the lines in $\mathcal{L}_1 \cup \mathcal{L}_2 \cup \cdots \cup \mathcal{L}_i$ and let $\mathcal{B}$ be a subset of $\mathcal{S}$. Let E be the event that the lines infected in the first i epochs are precisely those in $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_i$; also let E' denote the event that set of boosts in $\mathcal{S}$ is precisely $\mathcal{B}$. Since $\mathcal{E}$ is the disjoint union of such events E, it suffices to show that $\mathbb{P}(\mathcal{F} \,|\, E) = O((n(np)^{r-h})^{v'})$ for every such event E. We shall demonstrate the stronger assertion that $\mathbb{P}(\mathcal{F} \,|\, E \cap E') = O((n(np)^{r-h})^{v'})$ for every pair of events E and E' as above.

Let $\mathcal{V}_j$ be the set of those vertical lines of $\mathcal{S}$ which meet $\mathcal{B}$ in j points and let $N_j = |\mathcal{V}_j|$.

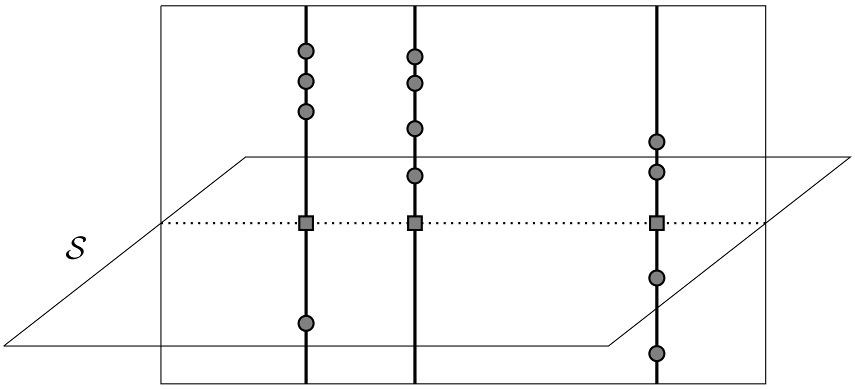

FIGURE 3. Boosts on a line in $\mathcal{S}$.

By conditioning on E', we are assuming that there are at least N_j planes with j or more parallel fully infected lines before the $(i+1)$-th epoch. Therefore, if $N_j > n^{1-j\gamma/(r-\gamma)}$ for some $1 \leq j \leq s+1$, then this implies that the modified three-dimensional process gets stopped before the $(i+1)$-th epoch and so we trivially have $\mathbb{P}(\mathcal{F} \mid E \cap E') = 0$ in this case. Hence, we may assume that $N_j \leq n^{1-j\gamma/(r-\gamma)}$ for each $1 \leq j \leq s+1$. Observe that $(s+1)\gamma/(r-\gamma) > 1$ since $(s+1)(s+2) > r$ and so $N_j = 0$ for $j \geq s+1$ since $n^{1-(s+1)\gamma/(r-\gamma)} < 1$.

Let v_j be the number of lines of $\mathcal{V}_j$ whose intersection with $\mathcal{S} \setminus \mathcal{B}$ contains $r-h-j$ or more initially infected points. Let F be the event that $\sum_{j=0}^{s} v_j \geq v'$. Note that a point of $\mathcal{S}$ which is infected before the $(i+1)$-th epoch is either initially infected or belongs to $\mathcal{B}$. Hence, $\mathbb{P}(\mathcal{F} \mid E \cap E') = \mathbb{P}(F \mid E \cap E')$. Note that F is an increasing event that depends only on the points in $\mathcal{S} \setminus \mathcal{B}$.

We claim that $E \cap E'$ is decreasing on $\mathcal{S} \setminus \mathcal{B}$. Let $\omega \in \{0,1\}^{[n]^3}$ be an initial configuration that belongs to $E \cap E'$ and let $\omega' \in \{0,1\}^{[n]^3}$ be such that $\omega'_x \leq \omega_x$ for $x \in \mathcal{S} \setminus \mathcal{B}$ and $\omega'_x = \omega_x$ for $x \notin \mathcal{S} \setminus \mathcal{B}$. We now check that $\omega' \in E \cap E'$. As before, we first note that since the percolation process is monotone and $\omega' \leq \omega$, the set of fully infected lines (not just in $\mathcal{P}$, but rather in all of $[n]^3$) at any stage when we start from ω' is a subset of the set of fully infected lines at that stage when we start from ω. Next, note that since $\omega \in E \cap E'$, when we start the line percolation process from ω, none of the lines through any of the points in $\mathcal{S} \setminus \mathcal{B}$ are infected before the $(i+1)$-th epoch; in other words, none of these points participate in the spread of infection before the $(i+1)$-th epoch. Therefore, since $\omega'_x = \omega_x$ for all $x \notin \mathcal{S} \setminus \mathcal{B}$, it follows that the set of fully infected lines at any stage before the $(i+1)$-th epoch when we start from ω' is actually the same as the set of fully infected lines at that stage when we start from ω. Thus, $\omega' \in E \cap E'$ and it follows that $E \cap E'$ is decreasing on $\mathcal{S} \setminus \mathcal{B}$.

It follows from Proposition 3.4 that $\mathbb{P}(F \mid E \cap E') \leq \mathbb{P}(F)$. Therefore, it suffices to show that $\mathbb{P}(F) \leq O((n(np)^{r-h})^{v'})$.

We first consider $\mathbb{P}(v_0 \geq l)$. The probability that there exist l or more vertical lines in $\mathcal{V}_0$ which meet $\mathcal{S} \setminus \mathcal{B}$ in $r-h$ initially infected points is

$$O\left(\binom{N_0}{l}\left((np)^{r-h}\right)^l\right) = O\left(\left(n(np)^{r-h}\right)^l\right)$$

since $N_0 \leq n$. On the other hand, for $1 \leq j \leq s$, we have

$$\begin{aligned}
\mathbb{P}(v_j \geq l) &= O\left(\binom{N_j}{l}\left((np)^{r-h-j}\right)^l\right) \\
&= O\left(\left(n(np)^{r-h}\right)^l \left(n^{1-\gamma/(r-\gamma)}p\right)^{-jl}\right) \\
&= O\left(\left(n(np)^{r-h}\right)^l\right)
\end{aligned}$$

since $N_j \leq n^{1-j\gamma/(r-\gamma)}$ and $n^{1-\gamma/(r-\gamma)}p = Cn^{-(\gamma+1)/(r-\gamma)} = o(1)$ as $\gamma \geq 0$. It follows that

$$\mathbb{P}(F) = \mathbb{P}\left(\sum_{j=0}^{s} v_j \geq v'\right) \leq \sum_{x_0,x_1,\ldots,x_s} \mathbb{P}(v_0 \geq x_0, v_1 \geq x_1, \ldots, v_s \geq x_s),$$

where the sum above is over all non-negative integer solutions to the equation $x_0 + x_1 + \cdots + x_s = v'$. First, note that the number of such solutions is at most $(v'+1)^s$. Next, note that since the sets $\mathcal{V}_j$ are disjoint, we have

$$\mathbb{P}(v_0 \geq x_0, v_1 \geq x_1, \ldots, v_s \geq x_s) = \prod_{j=0}^{s} \mathbb{P}(v_i \geq x_i) = O\left(\left(n(np)^{r-h}\right)^{\sum_{j=0}^{s} x_j}\right).$$

It follows that $\mathbb{P}(F) \leq O((n(np)^{r-h})^{v'})$ and this completes the proof of Lemma 5.2. $\square$

Given a preface $\boldsymbol{\ell}^* = ((l_i, d_i))_{i=0}^m$, we shall mimic the proof of Theorem 2.1 to bound $\mathbb{P}(E_k(\boldsymbol{\ell}^*))$. Let us write $h = \sum_{i \le m : d_i = e_0} l_i$ and $v = \sum_{i \le m : d_i = e_1} l_i$; note that $h, v < k$ since $\boldsymbol{\ell}^*$ is a preface. Recall that s is the greatest natural number such that $s(s+1) \le r$ and that $p = Cn^{-1-1/(r-\gamma)}$ satisfies $n(np)^{r-s} \ll 1$ and $n(np)^{r-s-1} \gg 1$. Hence, by applying Lemma 5.2 repeatedly, we see that the probability of $E_k(\boldsymbol{\ell}^*)$, up to constant factors, is bounded above by

$$(n(np)^r)^{l_0} \times \cdots \times \Big(n(np)^{r - \sum_{j<i:d_j \neq d_i} l_j}\Big)^{l_i} \cdots$$
$$\cdots \times \Big(n(np)^{r - \sum_{j<m:d_j \neq d_m} l_j}\Big)^{l_m} \times \max\big((n(np)^{r-h})^{k-v}, (n(np)^{r-v})^{k-h}\big),$$

where the last term accounts for the two possible directions in which we might generate the k parallel fully infected lines on the last epoch. By collecting together consecutive terms that correspond to adding lines along the same direction, we may assume without loss of generality that the directions alternate in the expression above. Then, an algebraic simplification analogous to the one in the proof of Theorem 2.1 yields

$$\mathbb{P}(E_k(\boldsymbol{\ell}^*)) = O\Big(\max\Big(n^{k+h}(np)^{rk+rh-kh}, n^{k+v}(np)^{rk+rv-kv}\Big)\Big).$$

This implies that for any preface $\boldsymbol{\ell}^*$, we have

$$\begin{aligned}\mathbb{P}(E_k(\boldsymbol{\ell}^*)) &= O\Big(\max_{0 \le l < k} n^{k+l}(np)^{rk+rl-kl}\Big)\\ &= O\Big(\max_{0 \le l < k} n^k(np)^{rk}\Big(n(np)^{r-k}\Big)^l\Big). \qquad (13)\end{aligned}$$

When $k \le s$, we see that the estimate for the probability of $E_k(\boldsymbol{\ell}^*)$ in (13) is maximised by taking $l = 0$, from which we conclude that

$$\mathbb{P}(E_k) = O\Big((n(np)^r)^k\Big) = O\Big(C^{rk} n^{-\frac{k\gamma}{r-\gamma}}\Big).$$

On the other hand, when $k = s+1$, the estimate for the probability of $E_k(\boldsymbol{\ell}^*)$ in (13) is maximised by taking $l = s$, from which we conclude that

$$\mathbb{P}(E_{s+1}) = O\Big(n^{2s+1}(np)^{r(2s+1)-s(s+1)}\Big) = O\big(C^{r(2s+1)-s(s+1)} n^{-1}\big).$$

We have therefore established the following.

Corollary 5.3. *For $1 \le k \le s$, we have*

$$\mathbb{P}(E_k) = O\big(C^{rk} n^{-k\gamma/(r-\gamma)}\big),$$

and

$$\mathbb{P}(E_{s+1}) = O\big(C^{r(2s+1)-s(s+1)} n^{-1}\big). \qquad \square$$

We are now in a position to prove Lemma 5.1.

Proof of Lemma 5.1. Recall that E_A is the event that we stop modified three-dimensional process on account of the number of planes containing k parallel fully infected lines exceeding $n^{1-k\gamma/(r-\gamma)}$ for some $1 \le k \le s+1$.

From Corollary 5.3, we see that expected number of planes with k parallel fully infected lines when we stop the modified process in three dimensions is $O(C^{rk} n^{1-k\gamma/(r-\gamma)})$ when $1 \le k \le s$ and $O(C^{r(2s+1)-s(s+1)})$ when $k = s+1$. By Markov's inequality, the probability that the number of planes containing k parallel fully infected lines exceeds $n^{1-k\gamma/(r-\gamma)}$ is $O(C^{rk})$ when $1 \le k \le s$ and $O(C^{r(2s+1)-s(s+1)})$ when $k = s+1$ since $\lfloor n^{1-(s+1)\gamma/(r-\gamma)} \rfloor = 0$. Applying the union bound, we get

$$\mathbb{P}(E_A) = O\left(\sum_{1 \le k \le s} C^{(rk} + C^{r(2s+1)-s(s+1)} \right). \qquad \square$$

The required lower bound on p_c follows immediately from Lemma 5.1. The lemma implies that $\mathbb{P}(E_A) \to 0$ as $C \to 0$ for all $r \ge 2$. Hence, if C is less than a suitably small constant, the probability that the three-dimensional r-neighbour line percolation process with $p = Cn^{-1-1/(r-\gamma)}$ generates a plane with $s+1$ parallel fully infected lines before reaching termination is less than $1/2$. Consequently, the probability of percolation is also less than $1/2$. This implies that $p_c(n, r, 3) = \Omega(n^{-1-1/(r-\gamma)})$ as required. This completes the proof of Theorem 2.2. $\square$

6. Minimal percolating sets

In this section, we prove Theorem 2.3 which tells us the size of a minimal percolating set. We shall make use of the polynomial method which has had many unexpected applications in combinatorics; see [67] for a survey of many of these surprising applications. While linear algebraic techniques have previously been used to study bootstrap percolation processes (see [16]), we believe that this application of the polynomial method is new to the field.

Let us begin with two simple facts about polynomials.

Proposition 6.1. *If $A \subset \mathbb{R}^d$ is a finite set with $|A| < r^d$, then there exists a non-zero polynomial $P_A \in \mathbb{R}[x_1, x_2, \ldots, x_d]$ of degree at most $r - 1$ in each variable which vanishes on A.*

Proof. Let $V \subset \mathbb{R}[x_1, x_2, \ldots, x_d]$ be the vector space of real polynomials in d variables of degree at most $r - 1$ in each variable. The dimension of V is clearly r^d. Consider the evaluation map from V to $\mathbb{R}^{|A|}$ which sends a polynomial P to $(P(v))_{v \in A}$. Clearly, this map is linear. Since we assumed that $|A| < r^d$, this map has a non-trivial kernel. The existence of P_A follows. $\square$

The next proposition is from [2]; see Theorem 1.2.

Proposition 6.2. *Let $P = P(x_1, x_2, \ldots, x_d)$ be a polynomial in d variables over $\mathbb{R}$. Suppose that the degree of P as a polynomial in x_i is at most k_i for $1 \le i \le d$, and let $S_i \subset \mathbb{R}$ be a set of at least $k_i + 1$ distinct elements of $\mathbb{R}$. If $P(u_1, u_2, \ldots, u_d) = 0$ for every d-tuple $(u_1, u_2, \ldots, u_d) \in S_1 \times S_2 \times \cdots \times S_d$, then P is identically zero.* $\square$

We are now ready to prove Theorem 2.3.

Proof of Theorem 2.3. Suppose for the sake of contradiction that there is a set $A \subset [n]^d$ which percolates with $|A| < r^d$. We shall derive a contradiction using the polynomial method. Let $P_A \in \mathbb{R}[x_1, x_2, \ldots, x_d]$ be a polynomial of degree at most $r-1$ in each variable which vanishes on A; that P_A exists follows from Proposition 6.1. We shall use P_A to follow the spread of infection.

We claim that the polynomial P_A vanishes on $A^{(t)}$ for every $t \geq 0$. We prove this claim by induction on t. The claim is true when $t = 0$ since $A^{(0)} = A$. Now, assume P_A vanishes on $A^{(t)}$ and consider a line $\mathcal{L}$ which gets infected when going from $A^{(t)}$ to $A^{(t+1)}$. It must be the case that $|\mathcal{L} \cap A^{(t)}| \geq r$. Since P_A vanishes on $A^{(t)}$, the restriction of P_A to $\mathcal{L}$ vanishes on $\mathcal{L} \cap A^{(t)}$. If the direction of $\mathcal{L}$ is $i \in [d]$, then the restriction of P_A to $\mathcal{L}$ is a univariate polynomial in the variable x_i of degree at most $r-1$. Since a non-zero univariate polynomial of degree at most $r-1$ has at most $r-1$ roots, the restriction of P_A to $\mathcal{L}$ has to be identically zero. Consequently, P_A vanishes on $A^{(t+1)}$ and the claim is proved.

Since A percolates, we conclude that P_A vanishes on $[n]^d$. It follows from Proposition 6.2 that P_A is zero and we have a contradiction. This proves the theorem. $\square$

7. Concluding remarks

There remain many challenging and attractive open problems, chief amongst which is the determination of $p_c(n, r, d)$ for all $d, r \in \mathbb{N}$. To determine $p_c(n, r, 3)$, we used a careful estimate for $\vartheta_p(n, r, 2)$ which is valid for all $0 \leq p \leq 1$. This estimate for $\vartheta_p(n, r, 2)$ depends crucially on the fact that the two-dimensional process reaches termination in a constant (depending on r, but not on n) number of steps. We believe that to determine $p_c(n, r, 4)$, one will need to determine $\vartheta_p(n, r, 3)$ for all $0 \leq p \leq 1$ but since it is not at all obvious that the three-dimensional process reaches termination in a constant number of steps with high probability, we suspect different methods will be necessary.

Part 3

Extremal graph theory

CHAPTER 8

Disjoint induced subgraphs of the same order and size

Joint work with Béla Bollobás, Teeradej Kittipassorn and Alex Scott.

1. Introduction

Given a graph G, can we guarantee that G contains two large, vertex-disjoint copies of the same graph? It follows from Ramsey's theorem that any graph on n vertices contains two vertex-disjoint isomorphic induced subgraphs on $\Omega(\log n)$ vertices; by considering a random graph on n vertices, it is easy to check that this is also best possible up to constant factors.

What if, rather than asking for two isomorphic subgraphs, we ask for two subgraphs that are the same with respect to one or more graph parameters? Caro and Yuster [36] considered the question of finding two vertex-disjoint subgraphs of a given graph of the same order which induce the same number of edges. For a graph G, let $f(G)$ be the largest integer k such that there are two vertex-disjoint induced subgraphs of G each on k vertices, both inducing the same number of edges and let $f(n)$ be the minimum value of $f(G)$ taken over all graphs on n vertices. Trivially, $f(n) \leq \lfloor n/2 \rfloor$; also, as shown by Ben-Eliezer and Krivelevich [23], equality holds (with high probability) for the Erdős–Rényi random graphs $G(n,p)$ for all $0 \leq p \leq 1$.

There was a large gap between the best known upper and lower bounds for $f(n)$. From below, one can easily show using the pigeonhole principle that $f(n) = \Omega(n^{1/3})$. As observed by Caro and Yuster, it is possible to improve this to $f(n) = \Omega(n^{1/2})$ using a well known result of Lovász determining the chromatic number of Kneser graphs. By considering a carefully constructed disjoint union of cliques, each on an odd number of vertices, Caro and Yuster showed that $f(n) \le n/2 - \Omega(\log\log n)$.

As expected, one can say more about $f(G)$ when G belongs to certain special graph classes. For example, Axenovich, Martin and Ueckerdt [7] showed that $f(G) \ge \lceil n/2 \rceil - 1$ when G is a forest; this is clearly best possible. Indeed, it is possible to get quite close to the trivial upper bound of $n/2$ when we restrict our attention to sparse graphs. In their paper, Caro and Yuster showed, for any fixed $\alpha > 0$, that if G is a graph on n vertices, then $f(G) \ge n/2 - o(n)$ provided G has at most $n^{2-\alpha}$ edges (or non-edges). Axenovich, Martin and Ueckerdt [7] later showed that the same holds for graphs with at most $o(n^2/(\log n)^2)$ edges.

Our main aim in this chapter is to narrow considerably the gap between the best known upper and lower bounds for $f(n)$, and thereby answer a question of Caro and Yuster [36].

Theorem 1.1. *For every $\varepsilon > 0$, there exists a natural number $N = N(\varepsilon)$ such that for any graph G on $n > N$ vertices, $f(G) \ge n/2 - \varepsilon n$. Consequently,*

$$n/2 - o(n) \le f(n) \le n/2 - \Omega(\log\log n).$$

We remark that much research has been done on the family of induced subgraphs of a graph. For example, call a graph k-*universal* if it contains every graph of order k as an induced subgraph. Very crudely, if G is a k-universal graph with n vertices, then

$$\binom{n}{k} \ge \frac{2^{\binom{k}{2}}}{k!},$$

and so $n \ge 2^{(k-1)/2}$. As remarked in [32], almost all graphs with $k^2 2^{k/2}$ vertices are k-universal, and the Paley graphs come close to providing examples which are almost as good. Hajnal conjectured that if a graph only has a 'small' number of distinct (non-isomorphic) induced subgraphs, then it contains a trivial (complete or empty) subgraph with linearly many vertices. This was proved, shortly after the conjecture was made, by Alon and Bollobás [3], and Erdős and Hajnal [48], the latter in a stronger form. In [3] only a few parameters, like order, size and maximal degree, were used to distinguish non-isomorphic graphs.

Erdős and Hajnal [49] then went much further: they realised that forbidding a single graph as an induced subgraph severely constrains the structure of a graph. More precisely, they made the major conjecture that for every graph H, there is a positive constant $\gamma(H)$ such that if a graph of order n does not contain H as an induced subgraph, then the graph contains a trivial subgraph with at least $n^{\gamma(H)}$ vertices. In spite of all the work on this conjecture (for a small sample, see [38, 57, 94]) we are very far from the desired bound.

Let us finally mention another interesting line of research about finding disjoint isomorphic (not necessarily induced) subgraphs. Jacobson and Schönheim (see [50, 83]) independently raised the question of finding edge-disjoint isomorphic subgraphs. Improving on results of Erdős, Pyber and Pach [50], Lee, Loh and Sudakov [83] showed that every graph on m edges contains a pair of edge-disjoint isomorphic subgraphs with at least $\Omega((m \log m)^{2/3})$ edges and that this is best possible up to a multiplicative constant.

The rest of this chapter is organised as follows. We give an overview of our approach in Section 3, and then fill in the details and prove Theorem 1.1 in Section 4. There are many natural questions about induced subgraphs which are close to Theorem 1.1 in spirit; we conclude in Section 5 by mentioning some of these.

2. Preliminaries

Our objective in this section is to establish some notational conveniences and collect together, for easy reference, some simple propositions that we shall make use of when proving our main result.

2.1. Notation. A *pair* (x, y) will always mean an unordered pair with $x \neq y$, and a *collection* of pairs $\mathcal{P}$ will always mean a set of disjoint pairs; for example, $\mathcal{P} = \{(1, 2), (3, 4)\}$ is a collection of pairs, but $\mathcal{Q} = \{(1, 2), (2, 3)\}$ is not. For a collection of pairs denoted by $\mathcal{P}$, we shall write P for the underlying ground set of elements, i.e., $P = \bigcup_{(x,y)\in\mathcal{P}}\{x, y\}$; in other words, we reserve the corresponding upper case letter for the ground set. We shall say that two collections of pairs $\mathcal{P}$ and $\mathcal{Q}$ are disjoint if $P \cap Q = \varnothing$; for example, the collections $\mathcal{P}_1 = \{(1, 2), (3, 4)\}$ and $\mathcal{Q}_1 = \{(5, 6), (7, 8)\}$ are disjoint, while the collections $\mathcal{P}_2 = \{(1, 2), (3, 4)\}$ and $\mathcal{Q}_2 = \{(1, 3), (2, 4)\}$ are not.

As usual, given a graph $G = (V, E)$, we write $\deg(v)$ and $\Gamma(v)$ respectively for the degree and for the neighbourhood of a vertex v in G. For a subset $U \subset V$, we write $G[U]$ for the subgraph induced by U, $e(U)$ for the number of edges of $G[U]$, and $\deg(U)$ for the sum of the degrees (in G) of the vertices of U. Given two disjoint subsets $A, B \subset V$, we write $e(A, B)$ for the number of edges with one endpoint each in A and B.

We shall also use the following less common terminology and notation. For any two vertices $x, y \in V$, we write $\delta(x, y)$ for the *degree difference* between x and y, namely the quantity $|\deg(x) - \deg(y)|$. We say that two vertices x and y *disagree on a vertex* $v \neq x, y$ if v is adjacent to exactly one of x and y; otherwise x and y *agree* on v. For any two vertices $x, y \in V$, the *difference neighbourhood* $\Gamma(x, y)$ of x and y is the set of vertices $v \neq x, y$ on which x and y disagree; we write $\Delta(x, y)$ for the size of the difference neighbourhood, so that $\delta(x, y) \leq \Delta(x, y)$. If two vertices x and y agree on every vertex $v \neq x, y$, we say that the pair (x, y) is a *clone pair*. When the graph G in question is not clear from the context, we shall, for example, write $\delta(x, y, G)$ to denote the degree difference between x and y in G.

We say that a graph G is *splittable* if there is a partition $V = A \cup B$ of its vertex set into two sets A and B of equal size with $e(A) = e(B)$; in this case, we call (A, B) a *splitting* of G. Note that $e(A) = e(B)$ if and only if $\deg(A) = \deg(B)$, since $\deg(A) = 2e(A) + e(A, B)$.

Our conventions for asymptotic notation are largely standard; in particular, we we write $o_{k\to\infty}(1)$ to denote a function (of k) that goes to 0 as $k \to \infty$, and that when we write, say $\Omega_k(.)$, we mean that the constant suppressed by the asymptotic notation is allowed to depend on (but is completely determined by) the parameter k. For the sake of clarity of presentation, we systematically omit floors and ceilings whenever they are not crucial.

2.2. Preliminary observations. We shall make use of the following simple observation repeatedly when constructing a splitting.

PROPOSITION 2.1. *Given positive real numbers $x_1, x_2, \ldots, x_t$ in the interval $[a, b]$ with $0 \le a \le b$ and $y \in [-ta, ta]$, we may choose signs $\zeta_i \in \{-1, +1\}$ such that $|y + \sum \zeta_i x_i| \le b$.*

PROOF. We may assume without loss of generality that $y \ge 0$. Let j be the largest index such that $y > \sum_{i=1}^{j} x_i$; since $y \le ta$, we must have $j < t$. Set $\zeta_i = -1$ for $1 \le i \le j$. Now, clearly $y - \sum_{i=1}^{j} x_i \in [-b, b]$. Given a real number $z \in [-b, b]$ and a positive real $x \le b$, one of $z + x$ or $z - x$ always lies in $[-b, b]$. Consequently, we may choose the signs of $x_{j+1}, \ldots, x_t$ one-by-one, always ensuring that the partial sum is in the interval $[-b, b]$, thus proving the claim. $\square$

The following first moment bound will prove useful; it is easily checked that the bound is the best possible.

PROPOSITION 2.2. *Let X be a random variable such that $X \le N$ and $\mathbb{E}[X] \ge Np$. Then*

$$\mathbb{P}\left(X \ge \frac{\mathbb{E}[X]}{2}\right) \ge \frac{p}{2-p}.$$

PROOF. Let us write $t = \mathbb{P}(X \ge \mathbb{E}[X]/2)$. We know that $\mathbb{E}[X] \le tN + (1-t)\mathbb{E}[X]/2$. This implies that $t(N - \mathbb{E}[X]/2) \ge \mathbb{E}[X]/2$. The result follows from the fact that $\mathbb{E}[X] \ge Np$. $\square$

We will also need the following two easy propositions.

PROPOSITION 2.3. *Given $x_1, x_2, \ldots, x_t$ in the interval $[0, a]$, a positive real b and a natural number N, it is possible to find $\lfloor t/N \rfloor - \lceil a/b \rceil$ disjoint subsets* *of $\{x_1, x_2, \ldots, x_t\}$, each of size N, such that $|x_i - x_j| \le b$ for any x_i and x_j belonging to the same subset.*

Proof. Suppose that $x_1 \le x_2 \le \cdots \le x_t$. Let $i_0 = 1$ and define i_j to be the smallest index such that $x_{i_j} > x_{i_{j-1}} + b$ with the convention that if no such index exists, we set $i_j = t+1$ and stop. Consider the sets $S_j = \{x_{i_j}, x_{i_j+1}, \ldots, x_{i_{j+1}-1}\}$. Since $x_1 \ge 0$ and $x_t \le a$, there are at most $\lceil a/b \rceil$ such sets. Now, by discarding at most N numbers from each S_j if necessary, we can assume that N divides $|S_j|$ for each j. We now partition each S_j into subsets of size N. Clearly, $|x_i - x_j| \le b$ for any x_i and x_j belonging to the same subset. The number of elements we have discarded is at most $N\lceil a/b \rceil$. So the number of subsets of size N we are left with is at least $\lfloor t/N \rfloor - \lceil a/b \rceil$. $\square$

Remark. We shall often apply Proposition 2.3 to the degrees of a subset of vertices of a graph; we consequently obtain disjoint groups of vertices such that the degree difference of any two vertices in the same group is suitably bounded.

Proposition 2.4. *Let x, y and z be three vertices and U some subset of vertices of a graph G. Then some two of the vertices x, y and z disagree on at most two thirds of the vertices of U.*

Proof. Any vertex $v \in U$ belongs to at most two of the three difference neighbourhoods $\Gamma(x, y)$, $\Gamma(y, z)$ and $\Gamma(z, x)$. The claim follows by averaging. $\square$

2.3. Binomial random variables. We will need some easily proven statements about binomial random variables. We collect these here. As usual, for a random variable with distribution $\mathrm{Bin}(N, p)$, we write $\mu(= Np)$ for its mean and $\sigma^2(= Np(1-p))$ for its variance.

We begin by recalling the following proposition from the previous chapter.

Proposition 2.5. *Let X be a random variable with distribution $\mathrm{Bin}(N, p)$, with $p \le 1/2$. Then for any $1 \le k \le n$,*

$$\exp(-2\mu)(\mu/k)^k \le \mathbb{P}(X = k) \le \exp(-\mu)(2e\mu/k)^k.$$

Also, $\exp(-2\mu) \leq \mathbb{P}(X = 0) \leq \exp(-\mu)$. $\square$

We shall make use of the following standard concentration result which first appeared in a paper of Bernstein and was later rediscovered by Chernoff and Hoeffding; see Appendix A of [5] for example.

PROPOSITION 2.6. *Let X be a random variable with distribution* $\mathrm{Bin}(N, p)$. *Then*

$$\mathbb{P}(|X - Np| > t) \leq \exp\left(\frac{-t^2}{N/2 + 2t/3}\right). \qquad \square$$

PROPOSITION 2.7. *Let X be a random variable with distribution* $\mathrm{Bin}(N, p)$. *Then*

$$\mathbb{P}(X \textit{ is even}) = \frac{1}{2}(1 + (1 - 2p)^N). \qquad \square$$

To prove the next two propositions, it will help to define $\theta(N, p) = \max_k \mathbb{P}(X = k)$ where X has distribution $\mathrm{Bin}(N, p)$. It follows from Stirling's approximation, see [27], that $\theta(N, p) = O(1/\sqrt{p(1-p)N})$.

PROPOSITION 2.8. *Let X_1 and X_2 be two independent random variables both with distribution* $\mathrm{Bin}(N, p)$. *Then*

$$\mathbb{P}(X_1 = X_2) = o_{\sigma\to\infty}(1).$$

In particular, when $p \leq 1/2$, $\mathbb{P}(X_1 = X_2) = o_{\mu\to\infty}(1)$.

PROOF. The proof follows simply by noting that

$$\begin{aligned}\mathbb{P}(X_1 = X_2) &= \sum_k \mathbb{P}(X_1 = k)\mathbb{P}(X_2 = k) \\ &\leq \theta(N, p) \sum_k \mathbb{P}(X_2 = k) = \theta(N, p) = O(1/\sigma). \qquad \square\end{aligned}$$

PROPOSITION 2.9. *Let X_1 and X_2 be two independent random variables with distributions* $\mathrm{Bin}(N_1, p)$ *and* $\mathrm{Bin}(N_2, p)$ *respectively, with* $p \geq 1/2$. *Then*

$$\mathbb{P}(|X_1 - X_2| < |N_1 - N_2|^{1/3}) = o_{|N_1 - N_2|\to\infty}(1).$$

PROOF. Assume, without loss of generality, that $N_2 \geq N_1$ and write $t = N_2 - N_1$ and $q = 1 - p$. If $q \geq t^{4/5}/N_2$, then we proceed as follows. The probability that X_2 lies in an interval of size $2t^{1/3} + 1$ is, independently of the interval, at most

$$(2t^{1/3} + 1)\theta(N_2, p) = O\left(\frac{t^{1/3}}{\sqrt{pqN_2}}\right) = O(t^{-1/15}) = o(1);$$

the claim follows immediately in this case. Hence, suppose that $q < t^{4/5}/N_2$. In this case, we see from the standard Chernoff bound that $\mathbb{P}(X_2 < N_2 - 2t^{4/5}) = o(1)$. This implies that $\mathbb{P}(X_2 < N_1 + t^{1/3}) = o(1)$; since $X_1 \leq N_1$, the claim follows. □

PROPOSITION 2.10. *Let X_1 and X_2 be two independent random variables with distributions* $\mathrm{Bin}(N_1, p)$ *and* $\mathrm{Bin}(N_2, p)$ *respectively, with* $p \geq 1/2$. *Suppose* $N_1 \leq N$, $N_2 \leq N$ *and* $|N_1 - N_2| \leq cN^{1/2}$ *for some absolute constant* c. *Then*

$$\mathbb{P}(|X_1 - X_2| > N^{2/3}) = O\left(\exp\left(\frac{-N^{1/3}}{5}\right)\right).$$

PROOF. If $|X_1 - X_2| > N^{2/3}$, then since $|N_1 - N_2| \leq cN^{1/2}$, we must necessarily have either $|X_1 - \mathbb{E}[X_1]| \geq N^{2/3}/3$ or $|X_2 - \mathbb{E}[X_2]| \geq N^{2/3}/3$ assuming N is sufficiently large. By the Chernoff bound, we have

$$\mathbb{P}(|X_1 - \mathbb{E}[X_1]| \geq N^{2/3}/3) \leq \exp\left(\frac{-N^{4/3}/9}{N_1/2 + 2N^{2/3}/9}\right) \leq \exp\left(\frac{-N^{1/3}}{5}\right),$$

where the last inequality holds for all large enough N since $N_1 \leq N$. We have an analogous bound for $\mathbb{P}(|X_2 - \mathbb{E}[X_2]| \geq N^{2/3}/3)$; the claim follows. □

3. Overview of our strategy

To prove Theorem 1.1, we need to show that if $\varepsilon > 0$ and n is sufficiently large, then any graph G on n vertices contains two disjoint subsets of vertices of the same size, each of cardinality at least $(1/2 - \varepsilon)n$, which induce the same number of edges. Equivalently, we need to show that it is possible to transform

G into a splittable graph by deleting at most $2\varepsilon n$ vertices from G. Recall that a graph is splittable if and only if there is a partition of its vertex set into two sets of equal size such that the sums of the degrees of the vertices in the two sets are equal.

We shall show that there is a probability $0 < p \leq \varepsilon$ (depending on G) such that if we delete vertices from G with probability p, then the resulting graph H is splittable with positive probability.

To show that this random subgraph H is splittable, we shall exhibit a large collection of 'gadgets' in H. Given $0 \leq a \leq b$, by an $[a, b]$-*gadget*, we mean a pair of vertices (x, y) such that $a \leq \delta(x, y) \leq b$; a gadget, in other words, is just a pair of vertices whose degree difference we can control.

Once we have found sufficiently many suitable gadgets in H, we construct a splitting of H as follows: we use Proposition 2.1 to decide, one-by-one for each gadget, which way round to assign the vertices of the gadget to the sides of the splitting. The following lemma makes this idea precise.

LEMMA 3.1. *Let H be a graph on an even number of vertices and suppose that we can partition $V(H)$ into disjoint collections of pairs $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_k$ such that the pairs in $\mathcal{P}_i$ are $[a_i, b_i]$-gadgets, where $0 \leq a_1 \leq b_1$ and $0 < a_i \leq b_i$ for $2 \leq i \leq k$. If $b_{i-1} \leq a_i|\mathcal{P}_i|$ for each $2 \leq i \leq k$, then $V(H)$ can be partitioned into two sets A, B of the same size such that $|\deg(A) - \deg(B)| \leq b_k$. In particular, if $b_k = 1$, then H is splittable.*

PROOF. We show by induction on i that it is possible to partition the vertices of the gadgets in $\mathcal{P}_1, \dots, \mathcal{P}_i$ into two sets A_i and B_i of equal size such that $|\deg(A_i) - \deg(B_i)| \leq b_i$. The lemma follows by taking $A = A_k$ and $B = B_k$.

We set $b_0 = 0$ and $A_0 = B_0 = \varnothing$ and so the claim is trivially true when $i = 0$. So suppose that $i \geq 1$ and that we have constructed A_{i-1} and B_{i-1}. Denote the $[a_i, b_i]$-gadgets in $\mathcal{P}_i$ by (x_j, y_j), where $\deg(x_j) \geq \deg(y_j)$ for $1 \leq j \leq |\mathcal{P}_i|$. Using the fact that $b_{i-1} \leq a_i|\mathcal{P}_i|$, it follows from Proposition 2.1 that there is a choice of signs $\zeta_j \in \{-1, +1\}$ for $1 \le j \le |\mathcal{P}_i|$ such that

$$\left|(\deg(A_{i-1}) - \deg(B_{i-1})) + \sum_j \zeta_j \delta(x_j, y_j)\right| \le b_i.$$

Given ζ_j as above, we construct A_i and B_i from A_{i-1} and B_{i-1} as follows: for each $1 \le j \le |\mathcal{P}_i|$, we add x_j to A_{i-1} and y_j to B_{i-1} if $\zeta_j = 1$, and y_j to A_{i-1} and x_j to B_{i-1} if $\zeta_j = -1$. The claim follows.

If $b_k = 1$, notice that we have a partition of $V(H)$ into two sets A and B of equal size such that $|\deg(A) - \deg(B)| \le 1$. As $\deg(A) + \deg(B)$ is the sum of all the vertex degrees, we conclude that $\deg(A) = \deg(B)$ since $\deg(A) - \deg(B)$ must be even. $\square$

Lemma 3.1 tells us that a graph is splittable if we can find the right gadgets in the graph. The majority of the work in proving Theorem 1.1 is in showing that it is possible to find a good collection of gadgets.

4. Proof of the main result

We now try and make the intuition presented in Section 3 precise. We shall show that if $\varepsilon > 0$ and n is sufficiently large, it is possible to transform any graph G on n vertices into a splittable graph by deleting at most $2\varepsilon n$ vertices from G. Before we begin, we remark that the various constants suppressed by the asymptotic notation throughout the proof are allowed to depend on ε. We shall use $c_1, c_2, \ldots$ to represent small constants depending on ε and $C_1, C_2, \ldots$ for large constants depending on ε. All our estimates will hold when n is sufficiently large.

Proof of Theorem 1.1. Let $\varepsilon > 0$ be fixed. By deleting an arbitrary vertex of G if necessary, assume that $n = |V(G)|$ is even. Let $\beta = \beta(\varepsilon)$ be a small constant whose value we shall fix at the end of the argument in Case 1.

Call a pair of vertices (x, y) a 'large' pair if $\delta(x, y) \in [n^{1/3}, \beta n]$. Let $c_1 = \varepsilon/2$. We distinguish two cases depending on how many disjoint large pairs we can find in G. We first deal with the case when G contains many disjoint large pairs.

Case 1: G contains $c_1 n$ disjoint large pairs of vertices. In this case, we shall show that G either trivially has a large splittable induced subgraph or that G has an induced subgraph H of even order on at least $(1-2\varepsilon)n$ vertices that contains

(1) a collection $\mathcal{S}_H$ of $[1,1]$-gadgets of size $\Omega(n/\log n)$,
(2) a collection $\mathcal{M}_H$ of $[1,n^{2/3}]$-gadgets of size at least $2\beta n$, and
(3) a collection $\mathcal{L}_H$ of $[n^{1/9},2\beta n]$-gadgets of size $\Omega(n)$

such that the collections $\mathcal{S}_H$, $\mathcal{M}_H$, and $\mathcal{L}_H$ are disjoint. It is straightforward to check that such a graph H is splittable using Lemma 3.1. Indeed, pair up the vertices $V(H)\setminus(L_H\cup M_H\cup S_H)$ arbitrarily; any such pair is a $[0,n]$-gadget and so we have a partition of $V(H)$ into disjoint collections of $[0,n]$-gadgets, $[n^{1/9},2\beta n]$-gadgets, $[1,n^{2/3}]$-gadgets and $[1,1]$-gadgets. The sizes of these collections satisfy the conditions of Lemma 3.1 if n is sufficiently large and it follows that H is splittable.

We shall now show that G does indeed contain such an induced subgraph H. We shall construct H by deleting vertices from G at random.

To avoid notational clutter, in the rest of the argument in Case 1, we shall write *large-gadget* for an $[n^{1/9},2\beta n]$-gadget, *medium-gadget* for a $[1,n^{2/3}]$-gadget and *one-gadget* for a $[1,1]$-gadget.

Let $\mathcal{L}$ be a collection of $c_1 n$ large pairs of vertices of G. The pairs in $\mathcal{L}$ will be the candidates for the large-gadgets we hope to find in H. Our next task is to find a large collection $\mathcal{M}$ of 'medium' pairs and a reasonably large collection $\mathcal{S}$ of 'small' pairs; the collections $\mathcal{M}$ and $\mathcal{S}$ will provide the candidate pairs for the medium-gadgets and one-gadgets that we would like to find in H.

Now, $|V\setminus L| = (1-2c_1)n$; recall that in our notation, L denotes the underlying ground set of $\mathcal{L}$. If we find more than $(1/2-\varepsilon)n$ disjoint clone pairs (x,y) in $G[V\setminus L]$, we are done. Indeed, we can delete all the other $(\leq 2\varepsilon n)$

vertices not in any of these clone pairs to get a splittable graph: we split this graph by assigning different vertices of each clone pair to different halves of the partition. So we may assume that we can find a set $V' \subset V \setminus L$ of vertices of G such that any two vertices of V' disagree on some vertex of $V \setminus L$ and $|V'| \geq (2\varepsilon - 2c_1)n \geq \varepsilon n$.

Let $C_1 = 4/\varepsilon$ and let $c_2 = \varepsilon/12$. We now apply Proposition 2.3 to the degrees of the vertices of V'; by our choice of C_1 and c_2, we see that we can find $c_2 n$ disjoint groups of three vertices from V' such that $\delta(x, y) \leq C_1$ for any two vertices x and y in the same group. By Proposition 2.4, from each of these triples, we may choose a pair of vertices (x, y) such that $\Delta(x, y) \leq 2n/3$. Write $\mathcal{P}$ for this collection of $c_2 n$ pairs.

For $0 \leq i \leq \log n - 1$, let $\mathcal{P}_i$ be the collection of those pairs (x, y) in $\mathcal{P}$ such that $\Delta(x, y) \in [2^i, 2^{i+1})$. There are two possibilities that we need to consider. It might be that no collection $\mathcal{P}_i$ contains too many pairs; we deal with this case next. The case where one of these collections contains many pairs is easier; we deal with this scenario later with a modification of the argument that follows.

Let $C_2 \geq 4$ be a (large) constant depending on ε; we shall fix the value of C_2 later in the proof at the end of Case 1A. Also, let $c_3 = c_2/3C_2 \leq c_2/12$.

Case 1A: None of the collections $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n - 1}$ contains $c_3 n$ pairs. It is clear that at least one of the collections $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n-1}$ contains at least $c_2 n/\log n$ pairs. Let k be the smallest index such that $|\mathcal{P}_k| \geq c_3 n/\log n$ and let us define our collection of small pairs $\mathcal{S}$ by setting $\mathcal{S} = \mathcal{P}_k$. We now define our collection of medium pairs $\mathcal{M}$ by setting

$$\mathcal{M} = \mathcal{P}_{k+C_2} \cup \cdots \cup \mathcal{P}_{\log n - 1}.$$

Since k is minimal and $c_3 \leq c_2/12$, we see that $|\mathcal{M}| \geq c_2 n/2$.

We shall now restrict our attention to the collections $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$; note that they are disjoint. We shall make use of the following facts about these collections.

(1) $\mathcal{S}$ contains $c_3 n/\log n$ pairs of vertices (x, y) such that $\delta(x, y) \leq C_1$, $\Delta(x, y) \in [2^k, 2^{k+1})$, and $\Delta(x, y) \leq 2n/3$.

(2) $\mathcal{M}$ contains $c_2 n/2$ pairs of vertices (x, y) such that $\delta(x, y) \leq C_1$, and $\Delta(x, y) \geq 2^{k+C_2}$.

(3) $\mathcal{L}$ contains $c_1 n$ pairs of vertices (x, y) with $\delta(x, y) \in [n^{1/3}, \beta n]$.

(4) For any pair of vertices (x, y) in $\mathcal{S}$ or $\mathcal{M}$, there exists at least once vertex in $V \setminus L$ on which x and y disagree.

We are now in a position to describe how we intend to construct a splittable graph from G. We shall delete vertices from G independently with a fixed probability. We shall show that with positive probability, many of the small pairs from $\mathcal{S}$ form one-gadgets in the resulting graph, many of the medium pairs from $\mathcal{M}$ form medium-gadgets, and many of the large pairs from $\mathcal{L}$ form large-gadgets in the resulting graph.

Fix $p = \min\{\varepsilon, 2^{-k}\}$. We now delete vertices from G independently with probability p. Let H be the resulting graph. We shall show that with probability $\Omega(1)$, the graph H is splittable and contains at least $(1 - 2\varepsilon)n$ vertices; this clearly implies the result we are trying to prove.

Note that for a graph to be splittable, it must necessarily contain an even number of vertices. With this in mind, let $\mathcal{E}$ be the event that an even number of vertices have been deleted, in other words, $\mathcal{E}$ is the event that $|V(H)|$ is even. By Proposition 2.7, we see that $\mathbb{P}(\mathcal{E}) \geq 1/2$. We now analyse what happens to the degree differences of the pairs in $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$ in the graph H.

One-gadgets. We first show that many of the pairs in $\mathcal{S}$ form one-gadgets in H.

Lemma 4.1. *For any pair* $(x, y) \in \mathcal{S}$,

$$\mathbb{P}((x, y) \text{ is a one-gadget in } H \mid \mathcal{E}) \geq f(\varepsilon) > 0.$$

The crucial fact about Lemma 4.1 is that the lower bound on the probability is independent of C_2.

Proof of Lemma 4.1. Let $A = \Gamma(x) \setminus (\Gamma(y) \cup \{y\})$ and $B = \Gamma(y) \setminus (\Gamma(x) \cup \{x\})$. Thus, $\delta(x, y) = ||A| - |B||$ and $\Delta(x, y) = |A| + |B|$. Note that since x and y disagree on at least one vertex of $V \setminus L$, it cannot be the case that both A and B are empty. Suppose without loss of generality that $|A| \geq |B|$ and that in particular, $A \neq \varnothing$.

Let E_1 be the event that both x and y are not deleted, E_2 the event that no vertices are deleted from B, E_3 the event that exactly $|\delta(x, y) - 1|$ vertices are deleted from A, and E_4 the event that the number of vertices deleted from $V \setminus (A \cup B \cup \{x, y\})$ has the same parity as $|\delta(x, y) - 1|$. It is obvious that the family $\{E_1, E_2, E_3, E_4\}$ is independent since these events correspond to disjoint sets of vertices, and it is easy to check that

$$\begin{aligned} \mathbb{P}((x, y) \text{ is a 1-gadget in } H \mid \mathcal{E}) &\geq \mathbb{P}(\{(x, y) \text{ is a 1-gadget in } H\} \cap \mathcal{E}) \\ &\geq \prod_{i=1}^{4} \mathbb{P}(E_i). \end{aligned}$$

To complete our proof of the claim, we shall bound the factors on the right one by one. Clearly, $\mathbb{P}(E_1) \geq (1 - \varepsilon)^2$.

We trivially have $|A|, |B| \leq \Delta(x, y) < 2^{k+1}$. Furthermore $|A|, |B| \geq 2^{k-1} - C_1/2$, since $0 \leq \delta(x, y) \leq C_1$. Also, we know that $\varepsilon 2^{-k} \leq p \leq 2^{-k}$. To bound $\mathbb{P}(E_2)$, first note that $p|B| \leq 2$. Now, $\mathbb{P}(E_2) = \mathbb{P}(\mathrm{Bin}(|B|, p) = 0)$ and so, by Proposition 2.5, $\mathbb{P}(E_2) \geq \exp(-4)$.

We now bound $\mathbb{P}(E_3)$. Clearly, $p|A| \leq 2$. If $2^k \geq 2C_1$, then $|A| \geq 2^{k-2}$ and so $p|A| \geq \varepsilon/4$. If $2^k \leq 2C_1$, then $p \geq \varepsilon 2^{-k} \geq \varepsilon/2C_1$ and so $p|A| \geq \varepsilon/2C_1$ since $|A| \geq 1$. Consequently,

$$\min\{\varepsilon/4, \varepsilon/2C_1\} \leq p|A| \leq 2.$$

Now, $\mathbb{P}(E_3) = \mathbb{P}(\mathrm{Bin}(|A|, p) = |\delta(x, y) - 1|)$. Using the above estimates for $p|A|$ and the fact that $0 \leq \delta(x, y) \leq C_1$ in Proposition 2.5, we see that $\mathbb{P}(E_3) = \Omega_{\varepsilon, C_1}(1)$.

Finally, since $\Delta(x, y) \le 2n/3$, it follows that $|V \setminus (A \cup B)| \ge n/3$ and hence by Proposition 2.7, $\mathbb{P}(E_4) \ge 1/6$ for sufficiently large n. The claim follows. $\square$

From Lemma 4.1 and Proposition 2.2 we see that, conditional on $\mathcal{E}$, the number of one-gadgets in H from $\mathcal{S}$ is $\Omega(n/\log n)$ with probability at least $f(\varepsilon)/2$; furthermore, and crucially, we note that this lower bound on the probability is independent of the choice of C_2.

Medium-gadgets. We next shift our attention to the pairs in $\mathcal{M}$.

Lemma 4.2. *For any pair* $(x, y) \in \mathcal{M}$,

$$\mathbb{P}(1 \le \delta(x, y, H) \le n^{2/3} \,|\, x, y \in V(H)) = 1 - o_{C_2 \to \infty}(1) - o(1).$$

Proof. Let $N_1 = |\Gamma(x) \setminus (\Gamma(y) \cup \{y\})|$ and let $N_2 = |\Gamma(y) \setminus (\Gamma(x) \cup \{x\})|$ and suppose without loss of generality that $N_1 \ge N_2$. Note that $\delta(x, y) = |N_1 - N_2| \le C_1$. Let X_1 and X_2 be independent random variables with distributions $\mathrm{Bin}(N_1, 1-p)$ and $\mathrm{Bin}(N_2, 1-p)$ respectively. Observe that $\delta(x, y, H)$ has the same distribution as $|X_1 - X_2|$.

We condition on $x, y \in V(H)$. Let E_1 be the event that $\delta(x, y, H) = 0$. Clearly, $\mathbb{P}(E_1) = \mathbb{P}(X_1 = X_2)$. Let E_2 denote the event that $\delta(x, y, H) \ge n^{2/3}$. It is enough to show that $\mathbb{P}(E_1 \cup E_2) = o_{C_2 \to \infty}(1) + o(1)$.

For any fixed values of p and N_2, it is not hard to check that $\mathbb{P}(X_1 = X_2)$ attains its maximum when $N_1 = N_2$; indeed, to see this, note that $\mathbb{P}(X_1 = X_2) = \sum_{i=0}^{N_2} \mathbb{P}(X_1 = i)\mathbb{P}(X_2 = i)$ and the required conclusion follows from Cauchy–Schwarz inequality. Thus $\mathbb{P}(E_1)$ is bounded above by the probability of two independent random variables with the distribution $\mathrm{Bin}(N_2, 1-p)$, or equivalently $\mathrm{Bin}(N_2, p)$, being equal. Now, $N_2 \ge 2^{k+C_2-1} - C_1/2$ and $p \ge \varepsilon 2^{-k}$. Recall that $C_1 = 4/\varepsilon$; therefore, $pN_2 \ge \varepsilon 2^{C_2-1} - 2^{-k+1}$ which, since $k \ge 0$, means that $pN_2 \ge \varepsilon 2^{C_2-1} - 2$. As ε is fixed, we note that pN_2 can be made arbitrarily large by choosing C_2 large enough. Since $p \le 1/2$, by Proposition 2.8, we see that $\mathbb{P}(E_1) = o_{C_2 \to \infty}(1)$.

Clearly $\mathbb{P}(E_2) = \mathbb{P}(|X_1 - X_2| \geq n^{2/3})$. Applying Proposition 2.10 to X_1 and X_2, we conclude that $\mathbb{P}(E_2) = O(\exp(-n^{1/3}/5))$. $\square$

Let $\mathcal{M}'$ be the collection of those pairs $(x, y) \in \mathcal{M}$ such that both x and y survive in H. Since the family of events $\{x, y \in V(H)\}$ is a family of mutually independent events for different pairs $(x, y) \in \mathcal{M}$ and since $\mathbb{P}(x, y \in V(H)) \geq (1-\varepsilon)^2$, it follows from Proposition 2.6 that $\mathbb{P}(|\mathcal{M}'| < (1-\varepsilon)^2|\mathcal{M}|/2) = \exp(-\Omega(n))$.

From Lemma 4.2, it follows that for any pair $(x, y) \in \mathcal{M}$,

$$\mathbb{P}\Big(1 \leq \delta(x, y, H) \leq n^{2/3} \,\Big|\, (x, y) \in \mathcal{M}'\Big) = 1 - o_{C_2 \to \infty}(1) - o(1).$$

Thus, by Markov's inequality, the number of medium-gadgets in H from $\mathcal{M}'$ is at least $|\mathcal{M}'|/2$ with probability $1 - o_{C_2 \to \infty}(1) - o(1)$. Thus, the number of medium-gadgets in H is at least $(1-\varepsilon)^2|\mathcal{M}|/4$ with probability $1 - o_{C_2 \to \infty}(1) - o(1)$.

Thus, conditional on the event $\mathcal{E}$, the number of medium-gadgets in H from $\mathcal{M}$ is $\Omega(n)$ with probability $1 - o_{C_2 \to \infty}(1) - o(1)$.

Large-gadgets. We finally consider the pairs of vertices in $\mathcal{L}$. Recall that every pair $(x, y) \in \mathcal{L}$ is such that $\delta(x, y) \in [n^{1/3}, \beta n]$ where β is a (small) constant whose value we have yet to fix. (Indeed, the value of β has so far played no role in our calculations.)

Lemma 4.3. *For any pair $(x, y) \in \mathcal{L}$,*

$$\mathbb{P}(n^{1/9} \leq \delta(x, y, H) \leq 2\beta n \,|\, x, y \in V(H)) = 1 - o(1).$$

Proof. We condition on $x, y \in V(H)$. Let E_1 denote the event that $\delta(x, y, H) < n^{1/9}$. Since $\delta(x, y) \geq n^{1/3}$, it follows immediately from Proposition 2.9 that $\mathbb{P}(E_1) = o(1)$.

Let E_2 be the event that $\delta(x, y, H) > 2\beta n$. Let $A = \Gamma(x) \setminus (\Gamma(y) \cup \{y\})$ and $B = \Gamma(y) \setminus (\Gamma(x) \cup \{x\})$, and let X_1 and X_2 be random variables that denote the the number of vertices from A and B respectively which survive in H.

Clearly, the distributions of X_1 and X_2 are $\mathrm{Bin}(|A|, 1-p)$ and $\mathrm{Bin}(|B|, 1-p)$ respectively.

If E_2 were to occur, i.e., it were the case that $|X_1 - X_2| > 2\beta n$, then this would imply that either $|X_1 - (1-p)|A|| \geq \beta n/2$ or $|X_2 - (1-p)|B|| \geq \beta n/2$, since $(1-p)||A| - |B|| \leq \delta(x, y) \leq \beta n$. It follows that $\mathbb{P}(E_2) = o(1)$ since the probability of either of the above two possibilities is $\exp(-\Omega(n))$ by Proposition 2.6. □

Arguing as in the case of medium-gadgets, we see from Lemma 4.3 that conditional on the event $\mathcal{E}$, the number of large-gadgets in H from $\mathcal{L}$ is $\Omega(n)$ with probability $1 - o(1)$.

Constructing a splitting. We now have a reasonably clear picture of what the degree differences in H of the pairs of vertices in $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$ look like. In summary, conditional on $\mathcal{E}$, we have demonstrated that in H, we can find

(1) a collection $\mathcal{S}_H$ of $\Omega(n/\log n)$ one-gadgets with probability $f(\varepsilon)/2$,
(2) a collection $\mathcal{M}_H$ of $\Omega(n)$ medium-gadgets with probability $1 - o(1) - o_{C_2 \to \infty}(1)$, and
(3) a collection $\mathcal{L}_H$ of $\Omega(n)$ large-gadgets with probability $1 - o(1)$

such that the collections $\mathcal{S}_H, \mathcal{M}_H$ and $\mathcal{L}_H$ are disjoint.

Thus by choosing C_2 to be a sufficiently large constant depending on ε, by the union bound, we find all of the above with probability $\Omega(1)$ conditional on $\mathcal{E}$, provided n is sufficiently large. Also, the expected number of vertices deleted is at most εn and so by Proposition 2.6, the probability that we have deleted more than $2\varepsilon n$ vertices is $\exp(-\Omega(n))$.

Consequently, we see that H, with probability $\Omega(1)$, has the aforementioned collections of gadgets, and furthermore, also has an even number of vertices and at least $(1 - 2\varepsilon)n$ vertices. We are done if we can guarantee that $2\beta n \leq |\mathcal{M}_H|$; this is possible if we choose $\beta = \beta(\varepsilon)$ to be a suitably small constant because $|\mathcal{M}_H| = \Omega(n)$.

We now consider the case where one of the sets $\mathcal{P}_i$ contains many pairs.

Case 1B: One of the sets $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n-1}$ contains c_3n pairs. This case is easier to deal with than the previous one We shall argue exactly as before; however we shall have no need of medium-gadgets and it will suffice to consider one-gadgets and large-gadgets alone.

Let k be any index such that $|\mathcal{P}_k| \geq c_3n$ (while we chose k to be minimal previously, any index k such that $|\mathcal{P}_k| \geq c_3n$ will do in this case). As before, we set $p = \min\{\varepsilon, 2^{-k}\}$ and $\mathcal{S} = \mathcal{P}_k$. We now delete vertices from G independently with probability p. Let H be the resulting graph; as before, we condition on deleting an even number of vertices. We claim that H is splittable with probability $\Omega(1)$.

It is not hard to check that Lemma 4.1 and Lemma 4.3 hold in this case as well. We conclude that we can delete an even number of vertices from G to obtain a graph H with $|V(H)| \geq (1-2\varepsilon)n$ in such a way that in H, we can find

(1) a collection $\mathcal{S}_H$ of $\Omega(n)$ one-gadgets, and

(2) a collection $\mathcal{L}_H$ of $\Omega(n)$ large-gadgets

such that $\mathcal{S}_H$ and $\mathcal{L}_H$ are disjoint. As before, it follows from Lemma 3.1 that H is splittable when n is sufficiently large provided $2\beta n \leq |\mathcal{S}_H|$; this is possible if we choose $\beta = \beta(\varepsilon)$ to be a suitably small constant because $|\mathcal{S}_H| = \Omega(n)$.

Thus, for sufficiently small β (chosen so as to satisfy the conditions from both Case 1A and 1B), we see that we are done if G contains many disjoint large pairs. Note that we have now fixed the value of β. We now deal with the case G does not contain many disjoint large pairs.

Case 2: G does not contain c_1n disjoint large pairs. In this case, we shall show that G has an induced subgraph H of even order on at least $(1-2\varepsilon)n$ vertices such that $V(H)$ may be partitioned into

(1) a collection $\mathcal{S}_H$ of $[1,1]$-gadgets of size $\Omega(n/\log n)$, and

(2) a collection $\mathcal{M}_H$ of $[0, n^{2/3}]$-gadgets.

In the rest of the argument in Case 2, we shall, as before, call $[1,1]$-gadgets *one-gadgets* and we call $[0,n^{2/3}]$-gadgets (as opposed to $[1,n^{2/3}]$-gadgets as we did earlier) *medium-gadgets*.

It is easily seen from Lemma 3.1 that any graph H as above is splittable if n is sufficiently large. We construct our splitting by starting with the pairs in $\mathcal{M}_H$ - we can use these pairs to construct a partition such that sums of the degrees of the vertices of the two halves of the partition differ by at most $n^{2/3}$. We then use the the pairs in $\mathcal{S}_H$ to reduce the difference to at most one; we are done by parity considerations.

We now show how to find such a subgraph H. We start by describing how to find pairs of vertices which will be the candidates for the medium-gadgets we hope to find in H.

Let $\mathcal{L}$ be a maximal collection of large pairs in G. Note that since $\mathcal{L}$ is maximal, we have either $\delta(x,y) < n^{1/3}$ or $\delta(x,y) > \beta n$ for any two vertices $x,y \in V \setminus L$. As $\beta n > 2n^{1/3}$ for all sufficiently large n, there is a partition $V \setminus L = K_1 \cup K_2 \cup \cdots \cup K_m$ into 'clumps' K_i with $m \leq 1/\beta$ in such a way that $\delta(x,y) < n^{1/3}$ for any $x,y \in K_i$ and $\delta(x,y) > \beta n$ if $x \in K_i$ and $y \in K_j$ with $i \neq j$.

We ignore the way in which vertices are originally paired in $\mathcal{L}$ and focus on the ground set L. By Proposition 2.3, we can find from L, at least $|L|/2 - n^{1/2}$ disjoint pairs (x,y) such that $\delta(x,y) \leq n^{1/2}$; call this collection of pairs $\mathcal{Q}$.

Let F be the graph obtained from G as follows. Delete every vertex of $L \setminus Q$. Delete one vertex from every clump K which contains an odd number of vertices. Having done this, delete a clump K (i.e., delete all the vertices of K) if $|K| \leq n^{1/2}$.

Note that the vertex set of F consists of the surviving clumps, each of which has even size and cardinality at least $n^{1/2}$, and the (possibly empty) set of pairs $\mathcal{Q}$. Since we had at most $1/\beta$ clumps initially, we have deleted $O(n^{1/2})$ vertices in total from G to obtain F. Hence, for any two vertices $x,y \in V(F)$,

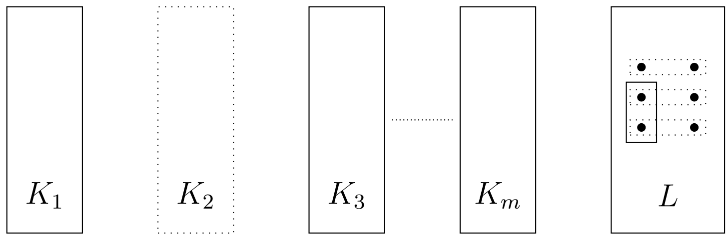

FIGURE 1. The graph F with small clumps removed and the vertices in L re-paired.

$|\delta(x,y,F) - \delta(x,y,G)| = O(n^{1/2})$. Hence, if either x and y both belong to the same (surviving) clump or if the pair (x,y) is in $\mathcal{Q}$, then $\delta(x,y,F) = O(n^{1/2})$. Let us say that two vertices $x, y \in V(F)$ are *proximate* if either both x and y belong to the same clump in F or if $(x,y) \in \mathcal{Q}$; these proximate pairs of vertices will be our candidates for medium-gadgets in H.

We now show how to find pairs of vertices which will be the candidates for the one-gadgets we hope to find in H. We shall henceforth work with F as opposed to G. We shall write V for $V(F)$ and all degrees and degree differences, unless specified otherwise, will be with respect to F.

Since $|\mathcal{L}| \leq c_1 n = \varepsilon n/2$ and since we have only deleted $O(n^{1/2})$ vertices so far, note that $|V \setminus Q| \geq (1 - 3\varepsilon/2)n$ for n sufficiently large.

If we find at least $(1/2 - \varepsilon)n$ disjoint clone pairs (x,y) in $F[V \setminus Q]$, this means that G trivially contains a large splittable induced subgraph we are done. So we may assume that we can find a set $V' \subset V \setminus Q$ of vertices of F with $|V'| \geq (2\varepsilon - 3\varepsilon/2)n = \varepsilon n/2$ such that any two vertices of V' disagree on some vertex in $V \setminus Q$.

We claim that if C_3 is sufficiently large (as a function of β), then we can find from any subset of C_3 vertices of V', two vertices x and y such that for each clump K, the number of vertices of K on which x and y disagree is at most $2|K|/3$. To see this, suppose that we have found C_3 vertices such that any two of them x and y disagree on more than two thirds of some clump $K_{x,y}$. Applying Ramsey's theorem (with $1/\beta$ colours) to the complete graph on these

C_3 vertices where the edge between x and y is labelled by the clump $K_{x,y}$, we see that we can find a monochromatic triangle provided C_3 is large enough. But by Proposition 2.4, out of any three vertices, at least two disagree on at most two thirds of the vertices of K. We have a contradiction.

Choose C_3 as described above and set $C_4 = 4C_3/\varepsilon$ and $c_4 = \beta/2C_4$. By Proposition 2.3, we can find from V', at least n/C_4 disjoint groups of size C_3 such that that $\delta(x, y) \leq C_4$ for any two vertices x and y in the same group. From each of these n/C_4 groups of size C_3, choose a pair of vertices (x, y) such that x and y disagree on at most two thirds of every clump. Choose a clump K^* such that at least a β fraction of these pairs (x, y) are such that x and y disagree on at least one vertex in K^*; this is possible because any two vertices of V' disagree on $V(F) \setminus Q$ and consequently, on at least one clump and furthermore, there are at most $1/\beta$ clumps. Let $\mathcal{P}$ be this collection of pairs which all disagree on at least one vertex in K^*; clearly $|\mathcal{P}| \geq \beta n/C_4 = 2c_4 n$.

We shall proceed as in Case 1 by pigeonholing the pairs in $\mathcal{P}$ into different boxes based on the size of their difference neighbourhoods, but with one important difference. Note that while *any two* vertices in the same clump have a small ($O(n^{1/2})$) degree difference, we can only guarantee that two vertices of Q have small ($O(n^{1/2})$) degree difference if the pair belongs to $\mathcal{Q}$. Consequently, when we later delete vertices at random, we shall either delete both vertices of a pair in $\mathcal{Q}$ or retain both; hence we shall treat a pair of vertices in $\mathcal{Q}$ as a single vertex when it comes to pigeonholing the pairs in $\mathcal{P}$. This is made precise below.

Let $F_{\mathcal{Q}}$ be the multigraph without loops obtained from F by contracting every pair (x, y) in $\mathcal{Q}$ (we ignore the loops that might arise). Note that there are at most two parallel edges between any two vertices of $F_{\mathcal{Q}}$. In $F_{\mathcal{Q}}$, we say that two vertices x and y disagree on a vertex $v \neq x, y$ if the number of edges between v and x is not equal to the number of edges between v and y. For $0 \leq i \leq \log n - 1$, let $\mathcal{P}_i$ be the collection of those pairs (x, y) in $\mathcal{P}$ such that

$\Delta(x, y, F_{\mathcal{Q}}) \in [2^i, 2^{i+1})$ where $\Delta(x, y, F_{\mathcal{Q}})$ is the number of vertices of $F_{\mathcal{Q}}$ on which x and y disagree.

As before, let k be any index such that $|\mathcal{P}_k| \geq 2c_4 n/\log n$; take $\mathcal{S} = \mathcal{P}_k$ and set $p = \min\{\varepsilon, 2^{-k}\}$.

In summary, $\mathcal{S}$ consists of pairs (x, y) such that

(1) x and y disagree on at most two thirds of every clump,

(2) x and y disagree on at least one vertex of K^*,

(3) $\delta(x, y) \leq C_4$, and

(4) $\Delta(x, y, F_{\mathcal{Q}}) \in [2^k, 2^{k+1})$.

Furthermore, since $\delta(x, y) \leq C_4 = o(n^{1/2})$ for any $(x, y) \in \mathcal{S}$, both members of any pair in $\mathcal{S}$ must belong to the same clump.

Consider the partition $\mathcal{S} = \mathcal{S}_o \cup \mathcal{S}_e$ where $\mathcal{S}_o$ is the set of those pairs $(x, y) \in \mathcal{S}$ such that $\delta(x, y)$ is odd. Recall that $|\mathcal{S}| \geq 2c_4 n/\log n$ and so one of $\mathcal{S}_o$ or $\mathcal{S}_e$ contains more than $c_4 n/\log n$ pairs. At this point, we need slightly different arguments depending on whether we have more pairs with odd degree difference or even degree difference in $\mathcal{S}$.

Case 2A: $\mathcal{S}$ contains many odd pairs. We first consider the case where $|\mathcal{S}_o| \geq c_4 n/\log n$. We shall delete vertices from F as follows. We pick vertices of $F_{\mathcal{Q}}$ independently with probability $p = \min\{\varepsilon, 2^{-k}\}$. For every vertex of $F_{\mathcal{Q}}$ that we pick, we delete (as appropriate) either the corresponding vertex or both vertices of the corresponding pair of vertices from $\mathcal{Q}$ in $F_{\mathcal{Q}}$. Let H be the resulting graph. Our aim is to show that H is splittable with probability $\Omega(1)$.

Earlier, we conditioned on deleting an even number of vertices from G. In this case, we need a little more. Let $\mathcal{E}^*$ be the event that an even number of vertices were deleted from each clump. By Proposition 2.7, we see that $\mathbb{P}(\mathcal{E}^*) \geq (1/2)^{1/\beta}$. Note that a consequence of $\mathcal{E}^*$ is that $|V(H)|$ is even.

One-gadgets. First, we shall show that many of the pairs in $\mathcal{S}_o$ become one-gadgets in H.

Lemma 4.4. *For any pair $(x, y) \in \mathcal{S}_o$,*

$$\mathbb{P}((x,y) \text{ is a one-gadget in } H \mid \mathcal{E}^*) = \Omega(1).$$

Proof. In $F_{\mathcal{Q}}$, let A be the set of those vertices $v \neq x, y$ such that number of edges between v and x is more than the number of edges between v and y and let B be defined analogously by interchanging x and y. Let $A = A_1 \cup A_2$ where A_1 and A_2 are respectively those vertices v in A such that the number of edges between v and x is one, respectively two, more than the number of edges from v to y; define B_1 and B_2 analogously.

The proof follows that of Lemma 4.1. Clearly,

$$2^k \leq |A_1| + |A_2| + |B_1| + |B_2| < 2^{k+1}.$$

Furthermore, $\delta(x, y) = ||A_1| + 2|A_2| - |B_1| - 2|B_2||$ and so,

$$-C_4 \leq |A_1| + 2|A_2| - |B_1| - 2|B_2| \leq C_4.$$

Using the above two inequalities, it is not hard to check that

$$\max\{|A_1|, |A_2|\}, \max\{|B_1|, |B_2|\} \geq 2^{k-3} - C_4/4.$$

Since $\delta(x, y)$ is odd, suppose without loss of generality that $\deg(x) > \deg(y)$. Let E_1 be the event that both x and y are not picked to be deleted, E_2 the event that no vertices are picked from B, E_3 the event that $X_1 + 2X_2 = \delta(x, y) - 1$ where X_1 and X_2 are the number of vertices picked from A_1 and A_2 respectively, and E_4 the event that the number of vertices picked from $K \setminus (A \cup B \cup \{x, y\})$ has the same parity as the number of vertices picked from $K \cap (A \cup B \cup \{x, y\})$ for every clump K. The collection of events $\{E_1, E_2, E_3\}$ is clearly independent, and it is easy to check that

$$\begin{aligned}
\mathbb{P}((x,y) \text{ is a 1-gadget in } H \mid \mathcal{E}^*) &\geq \mathbb{P}(\{(x,y) \text{ is a 1-gadget in } H\} \cap \mathcal{E}^*) \\
&\geq \mathbb{P}(E_1 \cap E_2 \cap E_3 \cap E_4) \\
&= \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3)\mathbb{P}(E_4 \mid E_1, E_2, E_3).
\end{aligned}$$

Clearly, $\mathbb{P}(E_1) \geq (1-\varepsilon)^2$. As in Lemma 4.1, note that $p|B| \leq 2$ and so, by Proposition 2.5, $\mathbb{P}(E_2) \geq \exp(-4)$.

We now bound $\mathbb{P}(E_3)$. First suppose that $2^{k-3} - C_4/4 > C_4$. Recall that $\delta(x,y)$ is odd. If $|A_2| \geq |A_1|$, we consider the event that $(\delta(x,y)-1)/2$ vertices are picked from A_2 and no vertices are picked from A_1 in $F_{\mathcal{Q}}$; as in Lemma 4.1, we see that $p|A_2| = \Theta(1)$ and so this event occurs with probability $\Omega(1)$. Hence E_3 occurs with probability $\Omega(1)$. If $|A_1| > |A_2|$, we consider the event that $\delta(x,y)-1$ vertices are picked from $|A_1|$ and no vertices are picked from A_2 and note that this event occurs with probability $\Omega(1)$ and hence E_3 occurs with probability $\Omega(1)$.

If, on the other hand, $2^{k-3} - C_4/4 \leq C_4$, then clearly $k = \Theta(1)$ and hence $p, |A_1|, |A_2|$ are all $\Theta(1)$. In this case, we consider the event that $t = \min\{(\delta(x,y)-1)/2, |A_2|\}$ vertices are picked from A_2 and $\delta(x,y) - 1 - 2t$ vertices are picked from A_1. Now, $|A_1| + 2|A_2| \geq \delta(x,y)$ since we assumed that $\deg(x) > \deg(y)$ and so $|A_1| \geq \delta(x,y) - 1 - 2t$. Also, as noted above, $p, |A_1|, |A_2|$ are all $\Theta(1)$. So this event occurs with probability $\Omega(1)$. Hence the event E_3 occurs with probability $\Omega(1)$.

Since x and y disagree on at most two thirds of every clump and since every clump has size at least $n^{1/2}$, it follows from Proposition 2.7 that for sufficiently large n, $\mathbb{P}(E_4 \mid E_1, E_2, E_3) \geq (1/6)^{1/\beta}$. $\square$

Let $\mathcal{S}_H$ be the set of pairs from $\mathcal{S}_o$ that form one-gadgets in H. From Lemma 4.4 and Proposition 2.2, we see that conditional on $\mathcal{E}^*$, $|\mathcal{S}_H| \geq \mathbb{E}[|\mathcal{S}_H|]/2 = \Omega(n/\log n)$ with probability $\Omega(1)$.

Medium-gadgets. We now show that the degree difference of any pair of vertices which are proximate in F cannot become too large in H.

Lemma 4.5. *Conditional on $\mathcal{E}^*$ and $|\mathcal{S}_H| \geq \mathbb{E}[|\mathcal{S}_H|]/2$, the probability that there exist $x, y \in V(H)$ which are proximate in F and satisfy $\delta(x,y,H) > n^{2/3}$ is $o(1)$.*

Proof. Recall that for any two vertices x and y which are proximate in F, $\delta(x,y) = O(n^{1/2})$. For such a pair of vertices x and y, note by Proposition 2.10 that

$$\mathbb{P}(\delta(x,y,H) > n^{2/3} \,|\, x,y \in V(H)) = O(\exp(-n^{1/3}/5)).$$

Consequently, since we have conditioned on an event with probability $\Omega(1)$, the probability that there exist some vertices $x, y \in V(H)$ such that x and y are proximate and $\delta(x,y,H) > n^{2/3}$ is $O(n^2 \exp(-n^{1/3}/5)) = o(1)$. $\square$

Constructing a splitting. We now describe how to construct a splitting of H. Let $\mathcal{Q}_H$ be the set of pairs from $\mathcal{Q}$ that survive in H. For a clump K in F, let K_H denote the set $(K \setminus S_H) \cap V(H)$. Clearly $V(H)$ is the disjoint union of S_H, Q_H and the clumps K_H. Note that conditional on $\mathcal{E}^*$, the size of K_H is even for every clump K since both members of any pair in $\mathcal{S}_H$ must necessarily belong to the same clump. Since each K_H has even cardinality, we may group the vertices of each K_H into pairs. Pair up the vertices in each K_H arbitrarily; let $\mathcal{M}_H$ be the collection consisting of these pairs and the pairs in $\mathcal{Q}_H$. Clearly, every pair of vertices in $\mathcal{M}_H$ are proximate in F and by Lemma 4.5, the probability that there exists some pair $(x,y) \in \mathcal{M}_H$ with $\delta(x,y,H) > n^{2/3}$ is $o(1)$.

The expected number of vertices deleted from F is at most εn and the number of vertices deleted from G to obtain F is $O(n^{1/2})$. Hence, by Proposition 2.6, the probability that we have deleted more than $2\varepsilon n$ vertices from G is $\exp(-\Omega(n))$.

We conclude that there exists an induced subgraph H of G such that $|V(H)| \geq (1-2\varepsilon)n$, and with the further property that $V(H)$ may be partitioned into

(1) a collection $\mathcal{S}_H$ of one-gadgets of size $\Omega(n/\log n)$, and
(2) a collection $\mathcal{M}_H$ of medium-gadgets.

It follows from Lemma 3.1 that H is splittable and we are done.

Case 2B: $\mathcal{S}$ contains many even pairs. Now we consider the case where $|\mathcal{S}_e| \geq c_4 n/\log n$.

Note that since we intend to delete either both vertices of a pair in $\mathcal{Q}$ or neither, it might be the case that it is impossible to make the parity of the degree difference of a pair in $\mathcal{S}_e$ odd in H. Consequently, in this case, we will need to work with $[2,2]$-gadgets, or *two-gadgets* for short, in addition to one-gadgets. With the exception of this slight change of tack to account for parity considerations, the argument is quite similar to the one in the previous case and proceeds as follows.

Let c_5 be a (small) constant depending on ε; the value of c_5 will be chosen later, following the statement of Lemma 4.6.

Recall that every pair of vertices in $\mathcal{S}_e$ disagree on some vertex in the clump K^*. Suppose there exists a vertex $v \in K^*$ such that $c_5 n/\log n$ pairs from $\mathcal{S}_e$ all disagree on v. In this case, we may complete the proof as follows. Let $\mathcal{S}_v \subset \mathcal{S}_e$ be the collection of pairs in $\mathcal{S}_e$ that disagree on v. We shall delete vertices from F as follows. We first delete v and then delete one other vertex uniformly at random from K^*. Following this, we proceed as before by picking vertices of $F_{\mathcal{Q}}$ independently with probability p and then deleting the corresponding vertices or pairs of vertices from $\mathcal{Q}$ in F. Let H be the resulting graph. Note that when we delete v, the degree difference of every pair in $\mathcal{S}_v$ changes parity and becomes odd. When we then delete another vertex uniformly at random from K^*, the parity of the degree difference of a pair in $\mathcal{S}_v$ is unaltered with probability at least $1/3$ since every pair in $\mathcal{S}$ disagree on at most two thirds of any clump. Arguing as in Lemma 4.4, for any pair in $\mathcal{S}_v$, we see that the probability that this pair forms a one-gadget in H, conditional on deleting an even number of vertices from every clump, is $\Omega(1)$ (albeit with a smaller constant than in Case 2A). Since $|\mathcal{S}_v| \geq c_5 n/\log n$, we can conclude the proof exactly as in the case where $\mathcal{S}$ contains many odd pairs.

Hence we may assume that for every vertex $v \in K^*$, the number of pairs in $\mathcal{S}_e$ that disagree on v is at most $c_5 n/\log n$. We delete vertices from F as before by picking vertices of $F_{\mathcal{Q}}$ independently with probability p and then deleting the corresponding vertices or pairs of vertices from $\mathcal{Q}$ in G. Let H be the resulting graph.

As before, let $\mathcal{E}^*$ be the event that an even number of vertices were deleted from each clump. The proof of Lemma 4.4, with minor modifications for the change in parity, yields a proof of the following lemma.

LEMMA 4.6. *For any* $(x, y) \in \mathcal{S}_e$, $\mathbb{P}((x, y)$ *is a two-gadget in* $H \mid \mathcal{E}^*) = \Omega(1)$. □

Let $\mathcal{S}_H$ be the collection of pairs from $\mathcal{S}_e$ that form two-gadgets in H. From Lemma 4.6 and Proposition 2.2, we see that there exists a small positive constant c_6 such that, conditional on $\mathcal{E}^*$, $|\mathcal{S}_H| \geq c_6 n/\log n$ with probability $\Omega(1)$. Recall that we still have not fixed the value of c_5; let us now fix $c_5 = c_6/4$.

CONSTRUCTING A SPLITTING. As before, let $\mathcal{Q}_H$ be the collection of pairs from $\mathcal{Q}$ that survive in H and for each clump K in F, let us write K_H for the set $(K \setminus S_H) \cap V(H)$.

We have shown that with probability $\Omega(1)$, the graph H is such that

(1) $|K_H|$ is even for every clump K, and
(2) $|\mathcal{S}_H| \geq c_6 n/\log n$.

Consider any pair $(x, y) \in \mathcal{S}_H$ and note that in F, x and y disagree on at most two thirds of any clump; in particular, x and y agree on at least a third of K^*. Consequently, the probability that x and y disagree on every vertex of K_H^* is $\exp(-\Omega(n^{1/2}))$. Hence, with probability $1 - o(1)$, for every $(x, y) \in \mathcal{S}_H$, there exists some vertex in K_H^* on which x and y agree.

Next, it follows from Lemma 4.5 that with probability $1 - o(1)$, any two vertices $x, y \in V(H)$ which are proximate satisfy $\delta(x, y, H) \leq n^{2/3}$. Finally, the probability that we have deleted more that $2\varepsilon n - 2$ vertices of total from G is, by Proposition 2.6, $\exp(-\Omega(n))$. It follows that with probability $\Omega(1)$, the graph H, in addition to possessing the aforementioned properties, also has the following properties.

(3) For every $(x, y) \in \mathcal{S}_H$, there exists some vertex in K_H^* on which x and y agree.

(4) For any $x, y \in V(H)$ such that x and y are proximate in F, $\delta(x, y, H) \leq n^{2/3}$.

(5) $|V(H)| \geq (1 - 2\varepsilon)n + 2$.

With a view to making the graph H splittable, we alter H as follows. Fix a pair $(x^*, y^*) \in \mathcal{S}_H$ and a vertex $v \in K^*$ on which x^* and y^* disagree. We know that there is a vertex $u \in K_H^*$ on which x^* and y^* agree. Delete u from H. If $v \in V(H)$, delete v from H and if $v \notin V(H)$, add v back. After these alterations, note that H still has an even number of vertices. Note also that now, $|V(H)| \geq (1 - 2\varepsilon)n$ and $\delta(x^*, y^*, H) \in \{1, 3\}$.

Before we altered H, at most $c_5 n / \log n$ pairs in $\mathcal{S}_H$ disagreed on any vertex in K^*; since $c_6 = 4c_5$, the alterations above change the degree differences of at most $2c_5 n / \log n = c_6 n / 2 \log n$ pairs in $\mathcal{S}_H$. Hence, H contains a collection $\mathcal{S}_H$ of least $c_6 n / 2 \log n - 1$ pairs of vertices (x, y) such that $\delta(x, y, H) = 2$ and a pair (x^*, y^*) such that $\delta(x^*, y^*, H) \in \{1, 3\}$. Furthermore, all the vertices of $V(H) \setminus (S_H \cup \{x^*, y^*\})$ may be grouped into pairs (x, y) such that $\delta(x, y, H) \leq n^{2/3} + 2$; let $\mathcal{M}_H$ denote this collection of pairs.

It is now easy to check that H is splittable using the argument used to prove Lemma 3.1. Indeed, we can use pairs in $\mathcal{M}_H$ to construct a partition such that sums of the degrees of the vertices of the two halves of the partition differ by at most $n^{2/3} + 2$. For n sufficiently large, we can then reduce the difference to at most two by using all but one of the pairs in $\mathcal{S}_H$. Finally, using the one remaining pair in $\mathcal{S}_H$ and the pair (x^*, y^*), we can reduce the difference to at most one; we are done constructing a splitting of H by parity considerations. This completes the proof of Theorem 1.1. $\square$

5. Concluding remarks

We have shown that $f(n) \geq n/2 - o(n)$. In fact, it should be possible to read out a bound of $f(n) \geq n/2 - n/(\log\log n)^c$ from our proof for some absolute constant $c > 0$; we chose not to include a proof of this fact to keep the presentation simple, and because we do not believe that such a bound is close to the truth. While we have managed to pin down f up to its first order term, there is still a large gap between the upper and lower bounds for $n/2 - f(n)$.

Problem 5.1. *What is the asymptotic behaviour of $n/2 - f(n)$?*

We know that $n/2 - f(n) = \Omega(\log\log n)$ and $n/2 - f(n) = o(n)$; we suspect that the truth lies closer to the lower bound and that in particular, $n/2 - f(n) = o(n^\varepsilon)$ for every $\varepsilon > 0$. Indeed, it is not inconceivable that $n/2 - f(n) = \Theta(\log n)$.

It is natural to generalise the problem to the case where we have more than one type of edge, or ask for more than two disjoint subgraphs. For any $r, l \in \mathbb{N}$, given an edge colouring Δ of the complete graph on n vertices with r colours, let $g(\Delta)$ be the largest integer k for which we can find l disjoint subsets $V_1, V_2, \ldots, V_l$ of $[n]$, each of cardinality k, such that for each $1 \leq i \leq r$, the number of edges induced by V_j of colour i is the same for every $1 \leq j \leq l$. Let $g(n, r, l)$ be the minimum value of $g(\Delta)$ taken over all edge colourings of the complete graph on n vertices. In particular, note that $g(n, 2, 2) = f(n)$. We conjecture that $g(n, r, 2) = n/2 - o(n)$ and more generally, ask the following question.

Problem 5.2. *For $r, l \in \mathbb{N}$, what is the asymptotic behaviour of $g(n, r, l)$?*

Finally, we mention a question about digraphs that we find particularly appealing. Given a digraph D on n vertices, let $h(D)$ denote the largest integer k for which there exist disjoint subsets $A, B \subset V$ such that $|A| = |B| = k$ and the number of directed edges from A to B is equal to the number of directed edges from B to A. Let $h(n)$ be the minimum value of $h(D)$ taken over all digraphs on n vertices.

PROBLEM 5.3. *Determine* $h(n)$.

CHAPTER 9

Catching a fast robber on the grid

Joint work with Paul Balister, Scott Binski and Béla Bollobás

1. Introduction

The game of *Cops and Robbers*, introduced almost thirty years ago independently by Nowakowski and Winkler [93] and Quilliot [95], is a perfect information pursuit-evasion game played on an undirected graph G as follows. There are two players, a set of cops and one robber. The game begins with the cops being placed onto vertices of their choice in G and then the robber, being fully aware of the placement of the cops, positions himself at a vertex of his choosing. Afterward, they move alternately, first the cops and then the robber along the edges of the graph G. In the cops' turn, each cop may move to an adjacent vertex, or remain where he is, and similarly for the robber; also, multiple cops are allowed to occupy the same vertex. The cops win if at some time there is a cop at the same vertex as the robber; otherwise, the robber wins. The minimum number of cops for which the cops have a winning strategy, no matter how the robber plays, is called the *cop number* of G.

Perhaps the most well known problem concerning the game of cops and robbers is Meyniel's conjecture which asserts that $O(\sqrt{n})$ cops are sufficient to catch the robber on any n-vertex graph. While Meyniel's conjecture has attracted a great deal of attention (see [9] and the references therein), progress towards the conjecture in its full generality has been rather slow.

In this chapter, we shall be concerned with a variant of the question where the robber is allowed to move faster than the cops. Let us suppose that the cops move normally as before while the robber is allowed to move at speed $R \in \mathbb{N}$; in other words, the robber may, on his turn, take any walk of length at most R from his current position that does not pass through a vertex occupied by a cop. The definition of the cop number in this setting is analogous. This variant was originally considered by Fomin, Golovach, Kratochvíl, Nisse and Suchan [54] and following them, Frieze, Krivelevich and Loh [60], Mehrabian [90], and Alon and Mehrabian [4] have obtained results about how large the cop number of an n-vertex graph can be when the robber has a fixed speed $R > 1$.

It is natural to ask how the cop number of a given graph changes, if at all, when the speed of the robber increases from 1 to some $R > 1$. The most natural example of a graph where this question is interesting is the $n \times n$ grid of squares where two squares of the grid are adjacent if and only if they share an edge. Let us write $f_R(n)$ for the minimum number of cops needed to catch a robber of speed R on an $n \times n$ grid. Maamoun and Meyniel [88] showed, amongst other things, that $f_1(n) = 2$ for all $n \geq 2$. However, the flavour of the problem changes completely as soon as the robber is allowed to move faster than the cops. Nisse and Suchan [92] showed that $f_2(n) = \Omega(\sqrt{\log n})$. Our aim in this chapter is to prove the following extension.

THEOREM 1.1. *There exists an $R \in \mathbb{N}$ and a $c_R > 0$ such that for all sufficiently large $n \in \mathbb{N}$,*

$$f_R(n) \geq \exp\left(\frac{c_R \log n}{\log \log n}\right).$$

To keep the presentation simple, we shall make no attempt to optimise the speed of the robber; we prove Theorem 1.1 with $R = 10^{25}$.

Note that $f_R(n) \leq n$ for every $R \in \mathbb{N}$ since n cops can catch a robber of any speed on the $n \times n$ grid by lining up on the bottom edge of the grid and then marching upwards together. We suspect that this trivial upper bound is closer to the truth than Theorem 1.1; we believe that for all sufficiently large $R \in \mathbb{N}$, $f_R(n) = n^{1-o(1)}$ as $n \to \infty$.

We give a sketch of the proof of Theorem 1.1 and then the proof proper in Section 2. We conclude with some discussion in Section 3.

2. Proof of the main result

Our proof of Theorem 1.1 is inspired by the strategy used by Bollobás and Leader [29] and Kutz [81] to resolve Conway's angel problem in three dimensions.

We fix a large positive integer $R \in \mathbb{N}$ which will denote the speed of the robber in what follows. We also fix two other positive integers $C, N \in \mathbb{N}$ such that C, N and R together satisfy $C \geq 40$, $N > 100e^C$ and $R > 50N$. We may, for example, take $C = 40$, $N = 10^{20}$ and $R = 10^{25}$.

Define a sequence of grids as follows: let $\mathcal{A}_0$ be an $N \times N$ grid and for $k \geq 1$ let $\mathcal{A}_k$ be a $(2k+1)^2 \times (2k+1)^2$ array of copies of $\mathcal{A}_{k-1}$.

We shall imagine that our $n \times n$ grid is tiled with copies of $\mathcal{A}_0$, with these copies of $\mathcal{A}_0$ themselves fitting together to form copies of $\mathcal{A}_1$, and so on. We call each copy of $\mathcal{A}_k$ in our grid a *k-cell*. Our strategy for the robber will be inductive: we shall describe how the robber may run from k-cell to adjacent k-cell, the path of the robber within a k-cell being inductively determined, all the while avoiding k-cells where there are too many cops.

Let us suppose that the robber is situated on the bottom edge of a 'safe' k-cell and wishes to get to the bottom edge of the k-cell above. Assume for the moment that the robber's k-cell is guaranteed to be 'safe' for a reasonably large number of steps.

Here then is an outline of a strategy for the robber: he plots a straight line from his current $(k-1)$-cell to a $(k-1)$-cell in the k-cell above that he wishes to get to. He runs across each of the $(k-1)$-cells on the way until he reaches his destination; within each $(k-1)$-cell, his path is determined inductively. Of course, there is a problem with this strategy: along the way, a $(k-1)$-cell that the robber needs to run across might not be 'safe' when he gets to it, or worse, a $(k-1)$-cell might become 'unsafe' while the robber is running through it. To address these issues, the robber alters his path dynamically and detours around any $(k-1)$-cell along his planned straight line path that he finds might become 'unsafe' while he is running through it. Our definition of 'safety' will ensure that the robber does not have to take too many detours. It will follow, and it is here that we use the fact that a k-cell is a $(2k+1)^2 \times (2k+1)^2$ array of $(k-1)$-cells, that the 'average speed' of the robber is large despite the fact that he has to take the occasional detour (and detours within detours and so on). This will provide us with enough elbow room to prove what we need by induction.

We now go about making the above sketch precise. First, we define a sequence $(L_k)_{k\geq 0}$ of natural numbers by setting $L_k = \prod_{j=0}^{k}(2j+1)^2$; clearly a k-cell is an $NL_k \times NL_k$ grid of squares. Next, we define another sequence $(T_k)_{k\geq 0}$ of natural numbers by setting $T_0 = 1$ and $T_k = (2k+1)^2 T_{k-1} + CT_{k-1}$ for $k \geq 1$.

An observation that we shall use repeatedly is that each $k \geq 0$,

$$T_k = L_k \prod_{j=1}^{k}\left(1 + \frac{C}{(2j+1)^2}\right) < L_k \exp\left(C\sum_{j=1}^{\infty}\frac{1}{(2j+1)^2}\right) < e^C L_k.$$

We need to define some notions of 'safety'. We say that a k-cell is *safe* at some point in time if the number of cops (at that point in time) within the k-cell is strictly less than 2^k. Also, we say that a k-cell is *safe for t steps* (at some point) if the set of cops at distance at most t from the k-cell has cardinality strictly less than 2^k. Note that a k-cell safe for t steps is necessarily safe for t' steps for every $0 \leq t' \leq t$ as well.

Next, we say that a square is *k-safe* if for each $0 \leq k' \leq k$, the k'-cell containing the square is safe for $T_{k'}$ steps; also, a square is *completely k-safe* if it is guaranteed to be k-safe after a single cop move.

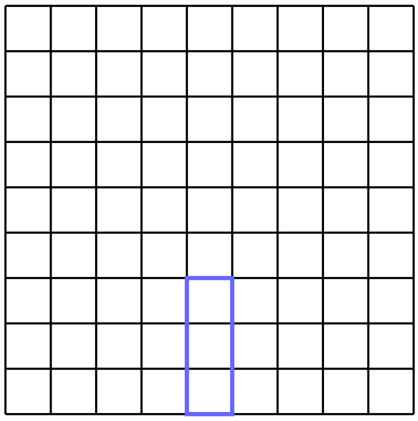

Figure 1. The bottom landing zone of a 1-cell; each square here represents a 0-cell.

Notice that if a k-cell is safe, then it contains at most one unsafe $(k-1)$-cell. We shall require a straightforward extension of this simple observation. Let us say that two cells are *separated* if they share neither an edge nor a corner.

Proposition 2.1. *Let X be a k-cell and assume that X is safe for t steps where $2t < NL_{k-1}$. If P and Q are a separated pair of $(k-1)$-cells within X, then either P or Q is safe for t steps.*

Proof. Simply notice that since P and Q are separated, the distance between them is at least NL_{k-1}. Consequently, the set of cops at distance at most t from P and the set of cops at distance at most t from Q are disjoint and the proposition follows. $\square$

To help with the induction, we shall demarcate certain regions as 'landing zones'. For $k \geq 1$, the *landing zone* of a k-cell is the union of its bottom, top, right and left landing zones; the *bottom landing zone* of a k-cell consists of the 3×1 sub-grid of $(k-1)$-cells at the middle of the bottom edge of the k-cell as shown in Figure 1 and the top, right and left landing zones are analogously defined by symmetry. Also, a square is called a *k-landing square*, if the square is contained in the landing zone of each k'-cell containing it for $1 \leq k' \leq k$.

Our proof of Theorem 1.1 hinges on the following lemma.

Lemma 2.2. *Let $k \geq 1$ and suppose that it is the robber's turn to move. Suppose further that the robber is positioned on a k-safe, k-landing square inside a k-cell X. If the k-cell Y above X is safe for $2T_k + 1$ steps, then the robber has a strategy to reach, in at most T_k steps and without getting caught, a k-landing square in the bottom landing zone of Y which is completely k-safe on his arrival there.*

Let us point out that there is some asymmetry in how Lemma 2.2 is stated. The lemma assumes something about the grid when the robber is about to move, and says something about the grid after a sequence of moves ending with the a move for the robber. However, note that a square is completely k-safe only if it is k-safe after a single cop move; hence, if the robber moves using the strategy given by Lemma 2.2, then no matter how the cops move on their turn following his final move, his new location is k-safe (and the lemma may be applied once again).

Proof of Lemma 2.2. Note that Lemma 2.2 is really a collection of four different statements, one each for when the robber starts in the bottom, top, right and left landing zones of his k-cell X. Indeed, Lemma 2.2 says that under certain conditions, it is possible for the robber to safely move from the landing zone of a k-cell to (the landing zone of) any of its four neighbouring k-cells in T_k steps.

We prove the lemma by induction on k. The case $k = 1$ is easy to check. Assume that it is the robber's turn to move, that he is on a 1-safe square in the landing zone of his 1-cell X, and that the 1-cell Y above him is safe for $2T_1 + 1 = 19 + 2C$ steps. We need to show that he can move in at most T_1 steps to a square in the bottom landing zone of Y which is completely 1-safe on his arrival. The robber can in fact do this in one step as we now describe.

Since the robber's square is 1-safe, note there are no cops in his 0-cell, say P. Consider a pair of separated 0-cells, call them Q and Q', in the bottom landing zone of Y. Note that at least one of Q or Q', say Q, must be safe for two steps because if not, then since $4 < NL_0 = N$, it follows from Proposition 2.1 that Y is not safe for two steps, contradicting our assumption that Y is safe for $19 + 2C$ steps with room to spare.

Note that a 1-cell is a 9×9 array of 0-cells. It is easy to check that there are N disjoint paths, each wholly contained within the union of X and Y and of length at most $36N$, from P to any 0-cell in Y. Since both X and Y are safe when the robber is about to move, there are at most two cops in total within X and Y. Hence, there are at least $N - 2$ paths between P and Q containing no cops on them. Note that the speed of the robber R is greater than $36N$, and hence the robber, on his turn, can follow one of these $N-2$ paths from his square in P to a square in Q. Note that Q is safe for two steps and Y is safe for $19 + 2C \geq T_1 + 1$ steps; hence, it clear that any square in Q is completely 1-safe on the robber's arrival there.

Now assume $k > 1$ and that we have proved the claim for each $1 \leq k' < k$. We describe the robber's strategy when he starts in the bottom landing zone of his k-cell. The strategy for the three other landing zones are very similar and we only highlight the very minor differences.

We shall divide the robber's journey into two parts. We first describe how the robber should travel from the bottom landing zone of his k-cell X to the top landing zone of X. This journey will require at most $(2k+1)^2 T_{k-1} + 3T_{k-1}$ steps. We then show that robber can dash across from the top landing zone of X into the bottom landing zone of Y in at most $13T_{k-1}$ steps. Hence, the total number of steps required will be bounded above by

$$(2k+1)^2 T_{k-1} + 16T_{k-1} \leq (2k+1)^2 T_{k-1} + CT_{k-1} \leq T_k$$

as required.

In what follows, when we speak of the robber arriving at a square in a $(k-1)$-cell, it is implied that the square is a $(k-1)$-landing square.

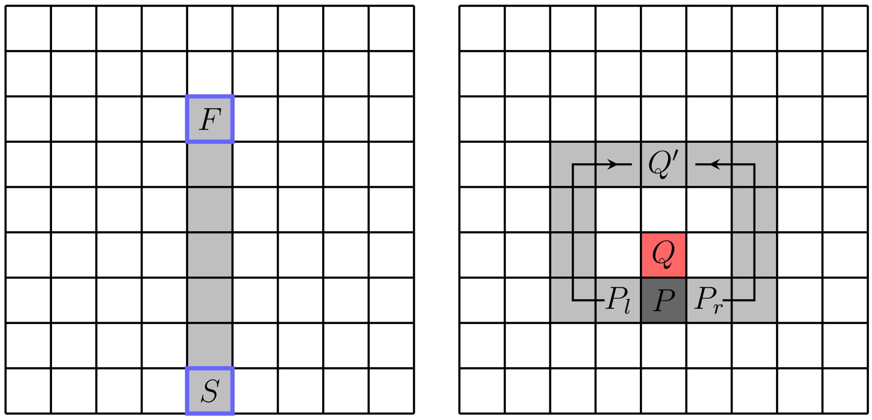

FIGURE 2. The planned path to the top landing zone and the detouring strategy.

The robber begins by plotting a straight line path from his $(k-1)$-cell, say S, to the nearest $(k-1)$-cell, say F, in the top landing zone of X as shown in Figure 2. The robber's square is k-safe; this means that X is safe for T_k steps and that the robber's square is $(k-1)$-safe. If the $(k-1)$-cell above him is safe for $2T_{k-1}+1$ steps, then the robber may inductively run, in at most T_{k-1} steps and without getting caught, to a square in the $(k-1)$-cell above him which is completely $(k-1)$-safe on his arrival there. Following the subsequent cop turn, his square is $(k-1)$-safe. The robber may repeat this process until he gets to F, provided that every time the robber arrives at a $(k-1)$-cell (and the cops have subsequently moved), the $(k-1)$-cell above is safe for $2T_{k-1}+1$ steps at that point. In this case, the robber reaches the top landing zone of X in at most $(2k+1)^2T_{k-1}$ steps, and we are done.

So we may assume that at some stage of his journey, the robber is on a $(k-1)$-safe square in a $(k-1)$-cell P within X, it is his turn to move, and that the $(k-1)$-cell Q above P is not safe for $2T_{k-1}+1$ steps. We claim that the robber only has to deal with such a situation once.

Let us consider the first time such a situation arises. Clearly, the robber has taken at most $(2k+1)^2T_{k-1}$ steps from S; as X was safe for $T_k = (2k+1)^2T_{k-1} + CT_{k-1}$ steps to begin with, X is now safe for at least CT_{k-1} steps. The robber

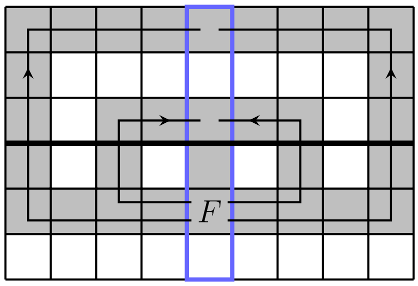

FIGURE 3. The final stretch.

takes a detour around Q as follows. He considers the two paths around Q to a $(k-1)$-cell Q' located above (and separated from) Q as shown in Figure 2; call these paths $\mathcal{Z}_l$ and $\mathcal{Z}_r$. We claim that each of the $(k-1)$-cells along one of these two paths is safe for $8T_{k-1}+1$ steps. Indeed, all the $(k-1)$-cells on these paths with the exception of the two initial $(k-1)$-cells P_l and P_r are separated from Q. If one of these $(k-1)$-cells is not safe for $8T_{k-1}+1$ steps, then since Q is not safe for $2T_{k-1}+1$ steps and

$$2(8T_{k-1}+1) \leq 18T_{k-1} < 18e^C L_{k-1} < NL_{k-1},$$

if follows by Proposition 2.1 that X is not safe for $8T_{k-1}+1 \leq 9T_{k-1}$ steps, contradicting the fact that X is in fact, safe for CT_{k-1} steps. Again, by Proposition 2.1, one of P_l and P_r is necessarily safe for $8T_{k-1}+1$ steps since P_l and P_r are separated. So suppose that all the $(k-1)$-cells along $\mathcal{Z}_l$ are safe for $8T_{k-1}+1$ steps. Then it is easy to check that the robber may inductively run along $\mathcal{Z}_l$, in at most $7T_{k-1}$ steps and without getting caught, from P to Q' so that he reaches a square in Q' which is completely $(k-1)$-safe on his arrival there.

Since Q was not safe for $2T_{k-1}+1$ steps when the robber was at P, we know that there are at least 2^{k-1} cops at distance at most $2T_{k-1}+1$ from Q; let us mark these cops. The robber takes $7T_{k-1}$ steps to reach Q' from P. In those $7T_{k-1}$ steps, the 2^{k-1} marked cops may move at most $7T_{k-1}$ steps up. However, since the distance between Q and Q' is $NL_{k-1} > 100e^C L_{k-1} > 100T_{k-1}$, it is clear that these 2^{k-1} marked cops can never overtake the robber vertically, and hence the robber will, after this detour, always find that when he arrives at a $(k-1)$-cell, the $(k-1)$-cell above him is safe for $2T_{k-1}+1$ steps.

It is therefore clear that the robber can safely reach some $(k-1)$-cell F in the top landing zone of X (though, on account of his detours, not necessarily his initial choice) in at most $(2k+1)^2T_{k-1}+3T_{k-1}$ steps. This completes the first leg of the robber's journey.

Let us now pause and survey the robber's situation after the cops have moved. He is now on a $(k-1)$-safe, $(k-1)$-landing square in a $(k-1)$-cell F in the top landing zone of his k-cell X. Also, X is now safe for at least $(C-3)T_{k-1}$ steps and Y, the k-cell above X, is safe for at least $T_k+(2C-3)T_{k-1}+1$ steps.

We now show that the robber can safely reach, in at most $13T_{k-1}$ steps, a square in the bottom landing zone of Y which is completely $(k-1)$-safe when the robber arrives there; that this square is also completely k-safe follows from the fact that Y is safe for at least $T_k+(2C-3)T_{k-1}+1$ steps before the robber starts the second leg of his journey.

Consider the set of four paths, as shown in Figure 3, from F to the $(k-1)$-cells in the bottom landing zone of Y. We claim that it follows from the fact that X is safe for $(C-3)T_{k-1}$ steps, and the fact that Y is safe for $T_k+(2C-3)T_{k-1}+1$ steps, that all the $(k-1)$-cells along one of these four paths are all safe for $14T_{k-1}+1$ steps. To see this, note that if a $(k-1)$-cell in X lying on one of the two paths to, say, the right of F is not safe for $14T_{k-1}+1$ steps, then we know from Proposition 2.1 that all the $(k-1)$-cells in X on both paths to the left of F are safe for $14T_{k-1}+1$ steps; the two paths to the left of F are completely separated within Y and hence, all the $(k-1)$-cells on one of these two paths must be safe for $14T_{k-1}+1$ steps. Since each of these four paths is composed of at most thirteen $(k-1)$-cells, it is clear that the robber can then complete his journey by following one of these paths in at most $13T_{k-1}$ steps. This completes the second leg of the robber's journey.

Clearly this also shows how the robber may proceed if he is initially located in the top landing zone of X. A similar strategy to the what has just been

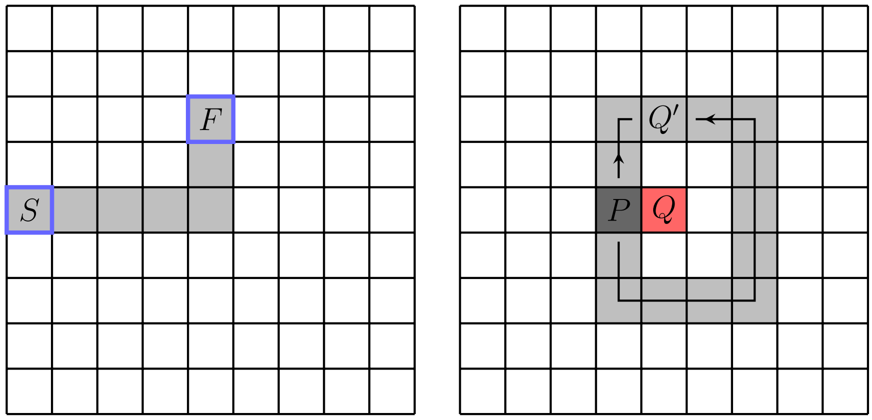

FIGURE 4. Getting to the top landing zone from the left landing zone.

described (see Figure 4) can be easily shown to work when the robber starts on either the right or the left landing zone of X; the robbers path becomes slightly longer than before if he needs to make a detour as he is 'turning', but our choices of C and N are large enough to ensure that the detouring strategy works with room to spare. $\square$

Armed with Lemma 2.2, it is a simple exercise to deduce Theorem 1.1.

PROOF OF THEOREM 1.1. We show that if $n \geq 2NL_k$ for some $k \geq 1$, then $f_R(n) \geq 2^k$. It is easy to check by estimating $L_k = \prod_{j=0}^{k}(2j+1)^2$ in terms of 2^k that this implies the result.

Since $n \geq 2NL_k$, we may fix a 2×2 array of k-cells in the grid. If the number of cops on the grid is strictly less than 2^k, each of these k-cells is guaranteed to be safe forever. After the cops have placed themselves on the grid, the robber positions himself on a k-safe, k-landing square in one of these four k-cells; that the robber can actually find such a square is easily checked by Proposition 2.1. The robber now wins by repeatedly using Lemma 2.2 to run around this 2×2 array in a clockwise loop forever. $\square$

3. Concluding remarks

As remarked earlier, it seems exceedingly unlikely that Theorem 1.1 is close to the truth; it should be the case that there exists an $R \in \mathbb{N}$ for which $f_R(n) = n^{1-o(1)}$ as $n \to \infty$.

Our proof of Theorem 1.1 is built on ideas used to solve Conway's angel problem in three dimensions. We conclude by mentioning that it is not inconceivable that one can, by suitably adapting one of the solutions (see [89, 33, 61]) to the angel problem in two dimensions, prove the existence of an $R \in \mathbb{N}$ and a $c_R > 0$ such that $f_R(n) \geq n^{c_R}$ for all sufficiently large $n \in \mathbb{N}$.

Bibliography

1. J. Adler and U. Lev, *Bootstrap Percolation: visualizations and applications*, Braz. J. Phys. **33** (2003), 641–644.
2. N. Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), 7–29.
3. N. Alon and B. Bollobás, *Graphs with a small number of distinct induced subgraphs*, Discrete Math. **75** (1989), 23–30.
4. N. Alon and A. Mehrabian, *On a generalization of Meyniel's conjecture on the Cops and Robbers game*, Electron. J. Combin. **18** (2011), Paper 19.
5. N. Alon and Joel H. Spencer, *The probabilistic method*, 3rd ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2008.
6. H. Amini, *Bootstrap percolation in living neural networks*, J. Stat. Phys. **141** (2010), 459–475.
7. M. Axenovich, R. R. Martin, and T. Ueckerdt, *Twins in graphs*, European J. Combin. **39** (2014), 188–197.
8. L. Babai, M. Simonovits, and J. Spencer, *Extremal subgraphs of random graphs*, J. Graph Theory **14** (1990), 599–622.
9. William Baird and Anthony Bonato, *Meyniel's conjecture on the cop number: a survey*, J. Comb. **3** (2012), 225–238.
10. P. Balister, S. Binski, B. Bollobás, and B. P. Narayanan, *Catching a fast robber on the grid*, Preprint.
11. P. Balister, B. Bollobás, J. Lee, and B. P. Narayanan, *Line percolation*, Random Structures Algorithms, To appear.
12. P. Balister, B. Bollobás, M. Przykucki, and P. Smith, *Subcritical neighbourhood family percolation models have non-trivial phase transitions*, Preprint,

arXiv:1311.5883.

13. J. Balogh, T. Bohman, and D. Mubayi, *Erdős-Ko-Rado in random hypergraphs*, Combin. Probab. Comput. **18** (2009), 629–646.
14. J. Balogh, B. Bollobás, H. Duminil-Copin, and R. Morris, *The sharp threshold for bootstrap percolation in all dimensions*, Trans. Amer. Math. Soc. **364** (2012), 2667–2701.
15. J. Balogh, B. Bollobás, and R. Morris, *Bootstrap percolation in three dimensions*, Ann. Probab. **37** (2009), 1329–1380.
16. J. Balogh, B. Bollobás, R. Morris, and O. Riordan, *Linear algebra and bootstrap percolation*, J. Combin. Theory, Ser. A **119** (2012), 1328–1335.
17. J. Balogh, B. Bollobás, and B. P. Narayanan, *Transference for the Erdős–Ko–Rado theorem*, Submitted to the Transactions of the American Mathematical Society.
18. J. Balogh, S. Das, M. Delcourt, H. Liu, and M. Sharifzadeh, *The typical structure of intersecting families of discrete structures*, Preprint, arXiv:1408.2559.
19. J. Balogh, R. Morris, and W. Samotij, *Independent sets in hypergraphs*, Preprint, arXiv:1204.6530.
20. ———, *Random sum-free subsets of Abelian groups*, Israel J. Math., To appear.
21. Zs. Baranyai, *On the factorization of the complete uniform hypergraph*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Colloq. Math. Soc. János Bolyai, vol. 10, North-Holland, Amsterdam, 1975, pp. 91–108.
22. M. Beiglböck, V. Bergelson, N. Hindman, and D. Strauss, *Some new results in multiplicative and additive Ramsey theory*, Trans. Amer. Math. Soc. **360** (2008), 819–847.
23. I. Ben-Eliezer and M. Krivelevich, *Perfectly balanced partitions of smoothed graphs*, Electron. J. Combin. **16** (2009), Note 14.

24. L. I. Bogolyubskiy, A. S. Gusev, M. M. Pyaderkin, and A. M. Raigorodskii, *The independence numbers and the chromatic numbers of random subgraphs of some distance graphs*, Mat. Sb., To appear.

25. ______, *The independence numbers and the chromatic numbers of random subgraphs of some distance graphs*, Doklady Math. **457** (2014), 383–387.

26. B. Bollobás, *On generalized graphs*, Acta Math. Acad. Sci. Hungar **16** (1965), 447–452.

27. ______, *Random graphs*, 2^{nd} ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.

28. B. Bollobás, T. Kittipassorn, B. P. Narayanan, and A. D. Scott, *Disjoint induced subgraphs of the same order and size*, European J. Combin., To appear.

29. B. Bollobás and I. Leader, *The angel and the devil in three dimensions*, J. Combin. Theory Ser. A **113** (2006), 176–184.

30. B. Bollobás, B. P. Narayanan, and A. M. Raigorodskii, *On the stability of the Erdős–Ko–Rado theorem*, Submitted to the Journal of Combinatorial Theory, Series A.

31. B. Bollobás, P. Smith, and A. Uzzell, *Monotone cellular automata in a random environment*, Combin. Probab. Comput., To appear.

32. B. Bollobás and A. Thomason, *Graphs which contain all small graphs*, European J. Combin. **2** (1981), 13–15.

33. B. H. Bowditch, *The angel game in the plane*, Combin. Probab. Comput. **16** (2007), 345–362.

34. G. Brightwell, K. Panagiotou, and A. Steger, *Extremal subgraphs of random graphs*, Random Structures Algorithms **41** (2012), 147–178.

35. N. Bushaw, M. Collares Neto, R. Morris, and P. Smith, *The sharp threshold for maximum size sum-free subsets in even-order Abelian groups*, Combin. Probab. Comput., To appear.

36. Y. Caro and R. Yuster, *Large disjoint subgraphs with the same order and size*, European J. Combin. **30** (2009), 813–821.

37. J. Chalupa, P. L. Leath, and G. R. Reich, *Bootstrap percolation on a Bethe lattice*, J. Phys. C. **12** (1979), 31–35.
38. M. Chudnovsky and S. Safra, *The Erdős-Hajnal conjecture for bull-free graphs*, J. Combin. Theory Ser. B **98** (2008), 1301–1310.
39. D. Conlon, *A new upper bound for diagonal Ramsey numbers*, Ann. of Math. **170** (2009), 941–960.
40. D. Conlon and T. Gowers, *Combinatorial theorems in sparse random sets*, Ann. of Math., To appear.
41. B. DeMarco and J. Kahn, *Mantel's theorem for random graphs*, Preprint, arXiv:1206.1016.
42. P. H. Diananda and H. P. Yap, *Maximal sum-free sets of elements of finite groups*, Proc. Japan Acad. **45** (1969), 1–5.
43. I. Dinur and E. Friedgut, *Intersecting families are essentially contained in juntas*, Combin. Probab. Comput. **18** (2009), 107–122.
44. H. Duminil-Copin and A. C. D. Van Enter, *Sharp metastability threshold for an anisotropic bootstrap percolation model*, Ann. Probab. **41** (2013), 1218–1242.
45. P. Erdős, *Some remarks on number theory*, Riveon Lematematika **9** (1955), 45–48.
46. P. Erdős, C. Ko, and R. Rado, *Intersection theorems for systems of finite sets*, Quart. J. Math. Oxford **12** (1961), 313–320.
47. P. Erdős and R. Rado, *A combinatorial theorem*, J. London Math. Soc. **25** (1950), 249–255.
48. P. Erdős and A. Hajnal, *On the number of distinct induced subgraphs of a graph*, Discrete Math. **75** (1989), 145–154.
49. ______, *Ramsey-type theorems*, Discrete Appl. Math. **25** (1989), 37–52.
50. P. Erdős, J. Pach, and L. Pyber, *Isomorphic subgraphs in a graph*, Combinatorics (Eger, 1987), Colloq. Math. Soc. János Bolyai, vol. 52, North-Holland, Amsterdam, 1988, pp. 553–556.

51. P. Erdős, M. Simonovits, and V. T. Sós, *Anti-Ramsey theorems*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), Colloq. Math. Soc. János Bolyai, vol. 10, North-Holland, Amsterdam, 1975, pp. 633–643.
52. M. Erickson, *A conjecture concerning Ramsey's theorem*, Discrete Math. **126** (1994), 395–398.
53. A. Fey, L. Levine, and Y. Peres, *Growth rates and explosions in sandpiles*, J. Stat. Phys. **138** (2010), 143–159.
54. F. V. Fomin, P. A. Golovach, J. Kratochvíl, N. Nisse, and K. Suchan, *Pursuing a fast robber on a graph*, Theoret. Comput. Sci. **411** (2010), 1167–1181.
55. L. R. Fontes, R. H. Schonmann, and V. Sidoravicius, *Stretched exponential fixation in stochastic Ising models at zero temperature*, Comm. Math. Phys. **228** (2002), 495–518.
56. K. Ford, *The distribution of integers with a divisor in a given interval*, Ann. of Math. **168** (2008), 367–433.
57. J. Fox and B. Sudakov, *Induced Ramsey-type theorems*, Adv. Math. **219** (2008), 1771–1800.
58. E. Friedgut, *On the measure of intersecting families, uniqueness and stability*, Combinatorica **28** (2008), 503–528.
59. E. Friedgut, V. Rödl, and M. Schacht, *Ramsey properties of random discrete structures*, Random Structures Algorithms **37** (2010), 407–436.
60. A. Frieze, M. Krivelevich, and P. Loh, *Variations on cops and robbers*, J. Graph Theory **69** (2012), 383–402.
61. P. Gacs, *The angel wins*, Preprint, arXiv:0706.2817.
62. M. M. Gauy, H. Hàn, and I. C. Oliveira, *Erdős-Ko-Rado for random hypergraphs: asymptotics and stability*, Preprint, arXiv:1409.3634.
63. R. L. Graham, B. L. Rothschild, and J. H. Spencer, *Ramsey theory*, 2nd ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., New York, 1990.

64. M. Granovetter, *Threshold models of collective behavior*, American J. Sociology **83** (1978), 1420–1443.

65. J. Gravner, C. Hoffman, J. Pfeiffer, and D. Sivakoff, *Bootstrap percolation on the hamming torus*, Ann. Appl. Probab., To appear.

66. B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. **167** (2008), 481–547.

67. L. Guth, *Unexpected applications of polynomials in combinatorics*, The Mathematics of Paul Erdős, Springer New York, 2013, pp. 493–522.

68. A. J. W. Hilton and E. C. Milner, *Some intersection theorems for systems of finite sets*, Quart. J. Math. Oxford **18** (1967), 369–384.

69. N. Hindman, I. Leader, and D. Strauss, *Forbidden distances in the rationals and the reals*, J. London Math. Soc. **73** (2006), 273–286.

70. A. E. Holroyd, *Sharp metastability threshold for two-dimensional bootstrap percolation*, Probab. Theory Related Fields **125** (2003), 195–224.

71. A. E. Holroyd, T. M. Liggett, and D. Romik, *Integrals, partitions, and cellular automata*, Trans. Amer. Math. Soc. **356** (2004), 3349–3368.

72. G. Katona, *A theorem of finite sets*, Theory of graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 187–207.

73. P. Keevash, *Shadows and intersections: stability and new proofs*, Adv. Math. **218** (2008), no. 5, 1685–1703.

74. P. Keevash and D. Mubayi, *Set systems without a simplex or a cluster*, Combinatorica **30** (2010), 175–200.

75. T. Kittipassorn and B. P. Narayanan, *Approximations to m-coloured complete infinite hypergraphs*, J. Graph Theory, To appear.

76. ______, *A canonical Ramsey theorem for exactly m-coloured complete subgraphs*, Combin. Probab. Comput. **23** (2014), 102–115.

77. D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), 167–172.

78. Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), 133–163.

79. Y. Kohayakawa, M. Schacht, and R. Spöhel, *Upper bounds on probability thresholds for asymmetric Ramsey properties*, Random Structures Algorithms **44** (2014), 1–28.
80. J. B. Kruskal, *The number of simplices in a complex*, Mathematical optimization techniques, Univ. of California Press, Berkeley, Calif., 1963, pp. 251–278.
81. M. Kutz, *Conway's angel in three dimensions*, Theoret. Comput. Sci. **349** (2005), 443–451.
82. I. Leader, P. A. Russell, and M. Walters, *Transitive sets in Euclidean Ramsey theory*, J. Combin. Theory Ser. A **119** (2012), 382–396.
83. C. Lee, P. Loh, and B. Sudakov, *Self-similarity of graphs*, SIAM J. Discrete Math. **27** (2013), 959–972.
84. I. H. Lee and A. Valentinyi, *Noisy contagion without mutation*, Rev. Econom. Stud. **67** (2000), 47–56.
85. L. Lovász, *On the Shannon capacity of a graph*, IEEE Trans. Inform. Theory **25** (1979), 1–7.
86. ______, *Combinatorial problems and exercises*, 2$^{\text{nd}}$ ed., AMS Chelsea Publishing, Providence, RI, 2007.
87. T. Łuczak, *Randomness and regularity*, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 899–909.
88. M. Maamoun and H. Meyniel, *On a game of policemen and robber*, Discrete Appl. Math. **17** (1987), 307–309.
89. A. Máthé, *The angel of power 2 wins*, Combin. Probab. Comput. **16** (2007), 363–374.
90. A. Mehrabian, *Lower bounds for the cop number when the robber is fast*, Combin. Probab. Comput. **20** (2011), 617–621.
91. B. P. Narayanan, *Exactly m-coloured complete infinite subgraphs*, J. Combin. Theory Ser. B **106** (2014), 163–173.
92. N. Nisse and K. Suchan, *Fast robber in planar graphs*, Graph-Theoretic Concepts in Computer Science, Lecture Notes in Computer Science, vol.

5344, Springer Berlin Heidelberg, 2008, pp. 312–323.

93. R. Nowakowski and P. Winkler, *Vertex-to-vertex pursuit in a graph*, Discrete Math. **43** (1983), 235–239.
94. H. J. Prömel and V. Rödl, *Non-Ramsey graphs are $c\log n$-universal*, J. Combin. Theory Ser. A **88** (1999), 379–384.
95. A Quilliot, *Jeux et pointes fixes sur les graphes*, Ph.D. thesis, Ph. D. Dissertation, Université de Paris VI, 1978.
96. F. P. Ramsey, *On a problem of formal logic*, Proc. London Math. Soc. **30** (1930), 264–286.
97. V. Rödl and A. Ruciński, *Threshold functions for Ramsey properties*, J. Amer. Math. Soc. **8** (1995), 917–942.
98. ______, *Rado partition theorem for random subsets of integers*, Proc. London Math. Soc. **74** (1997), 481–502.
99. ______, *Ramsey properties of random hypergraphs*, J. Combin. Theory Ser. A **81** (1998), 1–33.
100. V. Rödl and M. Schacht, *Extremal results in random graphs*, Preprint, arXiv:1302.2248.
101. K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.
102. D. Saxton and A. Thomason, *Hypergraph containers*, Preprint, arXiv:1204.6595.
103. M. Schacht, *Extremal results for random discrete structures*, Ann. of Math., To appear.
104. A. Stacey and P. Weidl, *The existence of exactly m-coloured complete subgraphs*, J. Combin. Theory Ser. B **75** (1999), 1–18.
105. G. Tenenbaum, *Sur la probabilité qu'un entier possède un diviseur dans un intervalle donné*, Compositio Math. **51** (1984), 243–263.
106. A. Thomason, *An upper bound for some Ramsey numbers*, J. Graph Theory **12** (1988), 509–517.

107. D. J. Watts, *A simple model of global cascades on random networks*, Proc. Nat. Acad. Sci. **99** (2002), 5766–5771.