On problems in Extremal Combinatorics

Manh Tuan Tran

Dissertation

zur Erlangung des Grades eines Doktors der Naturwissenschaften

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

Betreut durch

Prof. Tibor Szabó, PhD

Berlin, 2015

Erstgutachter: Prof. Tibor Szabó, PhD

Zweitgutachter: Prof. József Balogh, PhD

Tag der Disputation: 23.11.2015

Contents

List of Figures

To the memory of my grandmother (1919–2012)

Acknowledgement

First and foremost, I want to express my deepest gratitude to my advisor Tibor Szabó, for guiding me towards interesting and fruitful problems, for teaching me how to write papers and how to give talks, and for not giving up on me. To my mentor, Günter M. Ziegler, thank you for your time and precious advices.

This work would not have been possible without the support of the Research Training Group "Methods for Discrete Structures" (GRK 1408) and the Berlin Mathematical School. In particular I am thankful to Dorothea Kiefer for always answering my questions and helping me with various bureaucratic issues.

I would like to thank my co-authors: Józsi Balogh, Dennis Clemens, Shagnik Das, Julia Ehrenmüller, Hong Liu, Yury Person and Maryam Sharifzadeh for enjoyable times working together. I extend my gratitute as well to all of the people in Tibor's working group with whom I spent a wonderful time. This includes: Alexey, Anita, Chris, Christoph, Clément, Codrut, Dennis, Juanjo, Lothar, Shagnik, Yury.

I am grateful to Dennis and Chris for invaluable input on the summary of this thesis. For reading carefully some parts of my thesis and giving helpful remarks, let me thank Shagnik and Ky.

I am thankful to Józsi for inviting me to work with him and his students at UIUC, Hong and Maryam, to make my time in Illinois so fruitful and special.

I would like to express my warmest thanks to my friends Anh, Bum, Diep, Dũng, Huyen, Ky, Phi, Phuong, Quyen for filling my life with unforgettable moments spent with you.

My family has been encouraging and shown belief in me and my work. Their endless love and unconditioned support has been my source of inspiration, power and dedication.

Last but not least, I like to express my gratitude and deep feelings to Hà Phuong for enriching my life with her presence.

Chapter 1

Introduction

Extremal Combinatorics is one of the central branches of discrete mathematics and has developed spectacularly over the last few decades. It has lots of intriguing connections and important applications in various areas such as theoretical computer science, operations research, discrete geometry, probability theory and number theory.

Extremal Combinatorics studies how large or how small a structure can be, if it does not contain certain forbidden configuration. One of its major areas of study is extremal set theory, where the structures considered are families of sets, and the forbidden configurations are restricted intersection patterns. A fundamental result in this direction is the Erdős-Ko-Rado theorem [37] which determines the maximum size of uniform intersecting families. Another central area is extremal graph theory, in which the structures being studied are graphs, and the configurations to be avoided are given subgraphs. A basic result in this area is Turán's theorem [94], which gives the maximum number of edges in graphs with no copy of a given clique. Inspired from these theorems, countless extensions and variations have been developed. We shall discuss some of them in subsequent chapters.

One of the very active areas of research related to popular recreational games (e.g. Tic-Tac-Toe and Hex) is positional games. It enjoys fruitful interconnections with other combinatorial disciplines such as Ramsey theory, probabilistic combinatorics, and theoretical computer science. In the most general form, a positional game is a perfect information game described by a finite set of positions (the board) and by a family of subsets of the board (winning sets). Two players then alternatively claim previously unclaimed positions until they fully occupy the board. Different types of positional games are characterised by (different) rules that are used to determine the winner.

In this dissertation, we focus on various aspects of extremal combinatorics, including positional games, as well as the employment of the spectral method and the stability approach to study extremal problems. In the following, we shall briefly overview the topics that will be covered in the dissertation.

1.1 Removal and stability for Erdős-Ko-Rado

For positive integers n and k with $1 \leq k \leq n$, let $[n]$ denote the set of the first n natural numbers and let $\binom{[n]}{k}$ denote the family of all k-element subsets of $[n]$. A subfamily of $\binom{[n]}{k}$ is said to be *intersecting* if it does not contain a disjoint pair of sets. It is natural to ask how large such a family can be. When $n < 2k$, there are no two disjoint sets, and so $\binom{[n]}{k}$ is the maximal intersecting family. When $n \geq 2k$, a trivial construction is to take all sets containing some fixed element $i \in [n]$. This intersecting family contains $\binom{n-1}{k-1}$ sets which turns out to be the best possible due to Erdős, Ko and Rado [37].

A recent trend in extremal set theory is to go beyond the Erdős-Ko-Rado threshold and study the structure of families that need not to be intersecting, but contain few disjoint pairs. In this direction, Friedgut and Regev [49] proved a general removal lemma, showing that when $k = cn$ for some constant $0 < c < \frac{1}{2}$, a subfamily of $\binom{[n]}{k}$ with few disjoint pairs can be made intersecting by removing few sets. One of our main contributions in this chapter is to provide a simple proof of a special case of this theorem, when the family has size close to $\binom{n-1}{k-1}$. However, our theorem holds for all $2 \leq k < \frac{n}{2}$ and provides sharp quantitative estimates.

The *Kneser graph* $K(n, k)$ is the graph with vertex set $\binom{[n]}{k}$ such that two vertices are adjacent if they are disjoint. For $p = p(n, k) \in [0, 1]$, let $K_p(n, k)$ be the graph obtained from $K(n, k)$ by retaining each edge of $K(n, k)$ independently with probability p. Clearly, independent sets in $K(n, k)$ are nothing but intersecting families in $\binom{[n]}{k}$. Thus, the Erdős-Ko-Rado theorem can be restated saying the maximum size of independent sets in $K(n, k)$ is $\alpha(K(n, k)) = \binom{n-1}{k-1}$. As $K_p(n, k) \subset K(n, k)$, we must have $\alpha(K_p(n, k)) \geq \alpha(K(n, k)) = \binom{n-1}{k-1}$. Bollobás, Narayanan and Raigorodskii [15] asked to determine p for which the equality $\alpha(K_p(n, k)) = \binom{n-1}{k-1}$ holds. In this chapter we use our removal lemma to answer their question for $k = o(n)$, and provide strong bounds on the critical probability for $k \leq \frac{n-3}{2}$.
The results of this chapter is joint work with Shagnik Das [25].

1.2 Erdős-Rothschild problem for intersecting families

One of the fundamental results in graph theory is the theorem of Turán, proved in 1941, which initiated the development of Extremal Graph Theory. Basically, it states that the largest K_k-free graph on n vertices is $(k-1)$-partite. Thirty three years after the birth of Turán's theorem, Erdős and Rothschild [36] proposed a novel twist to the theorem: they asked for the maximum number of edge-colourings (not necessarily proper) of an n-vertex graph avoiding monochromatic copies of K_k, and wondered whether it would lead to extremal configurations that are substantially different from those of Turán's theorem. Substantial progress have been made on this problem over the past decade. For more detail on what has been done, we refer the reader to [84] and the references therein.

A problem with the same flavour has been addressed by Hoppen, Kohayakawa and Lefmann [62] in connection with the Erdős-Ko-Rado theorem. It can be stated as follows: given a set family $\mathcal{F}$, a (r,t)-colouring of $\mathcal{F}$ is a map from $\mathcal{F}$ to $[r]$ associating a colour with each element of $\mathcal{F}$ with the property that each colour class is a t-intersecting family. We write $c(\mathcal{F},r,t)$ for the number of (r,t)-colourings of $\mathcal{F}$, and let

$$c(n,k,r,t) = \max\{c(\mathcal{F},r,t) : \mathcal{F} \text{ is a } k\text{-uniform family on } [n]\}.$$

When k,r,t are fixed and n is sufficiently large, Hoppen, Kohayakawa and Lefmann determined the exact value of this function and the corresponding extremal families. In particular, they showed that when $r \in \{2,3\}$ and k,t are fixed, all extremal families are stars, thus implying stars are also the largest t-intersecting families. In Chapter 2, we allow k,r and t to grow as functions of n. We also address the problem in other settings, including permutations and vector spaces.
The results of this chapter is joint work with Dennis Clemens and Shagnik Das [20].

1.3 A Density Turán Theorem

Turán's theorem [94] states that every graph G of edge density $2e(G)/v(G)^2 > \frac{k-2}{k-1}$ contains a complete graph K_k and describes the unique extremal graph. The idea to study multipartite version of this theorem goes back to a suggestion by Bollobás (see the discussion after the proof of Theorem VI.2.15 in [12]). The first systematic investigations of this kind have been carried out by Bondy, Shen, Thomassé and Thomassen [17]. In the case of triangles they showed the following: let $d_\ell(K_3)$ denote the minimum real number with the property that any ℓ-partite graph G contains a triangle as soon as every edge density between two vertex classes of G is greater than $d_\ell(K_3)$. Then, $d_\ell(K_3)$ decreases to $\frac{1}{2}$ as ℓ tends to infinity. Bondy

et. al. also showed $d_3(K_3) = \frac{-1+\sqrt{5}}{2} \approx 0.61$, $d_4(K_3) > 0.51$, and speculated that $d_\ell(K_3) > \frac{1}{2}$ for all finite ℓ. Thus, it was a surprise when Pfender [85] managed to prove that actually $d_\ell(K_3) = \frac{1}{2}$ for all $\ell \geq 13$. He went on and showed that, for ℓ large enough, $d_\ell(K_k) = \frac{k-2}{k-1}$. For a general graph H, he suggested $d_\ell(H) = \frac{\chi(H)-2}{\chi(H)-1}$ for sufficiently large ℓ. In Chapter 4 we clarify the situation, showing Pfender's suggestion is not always true. In fact, we extend Pfender's idea to characterise all graphs H for which the equality $d_\ell(H) = \frac{\chi(H)-2}{\chi(H)-1}$ holds for $\ell \geq \ell_0(H)$ sufficiently large. The proof of our characterisation is an application of the stability method introduced by Simonovits [92].
The results of this chapter is joint work with Lothar Narins [81].

1.4 Keeping Avoider's graph almost acyclic

Let b be a positive integer and let $\mathcal{F} \subseteq 2^X$ be a hypergraph over a finite set X. In a *strict* $(1:b)$ Avoider-Enforcer game $\mathcal{F}$ two players, called *Avoider* and *Enforcer*, alternately occupy previously unoccupied elements of the so-called *board* X. Avoider occupies *exactly* 1 element per move and Enforcer occupies *exactly* b vertices per move. In a *monotone* $(1:b)$ Avoider-Enforcer game $\mathcal{F}$ in each turn Avoider claims *at least* 1 element of the board, where Enforcer claim *at least* b elements of the board. In both games, if the number of unclaimed elements is strictly less than b before a move of Enforcer, then he must occupy all of the remaining free vertices. The game ends when every element of the board has been claimed by one of the players. Avoider wins the game if he does not fully occupy a hyperedge of $\mathcal{F}$; otherwise Enforcer wins.

From now on, each game can be viewed under two different sets of rules – the *strict* game and the *monotone* game. Given a positional game $\mathcal{F}$, for its *strict* version we define the lower threshold bias $f_{\mathcal{F}}^-$ to be the largest integer such that Enforcer has a winning strategy for the $(1:b)$ game on $\mathcal{F}$ for every $b \leq f_{\mathcal{F}}^-$; and the upper threshold bias $f_{\mathcal{F}}^+$ to be the smallest non-negative integer such that Avoider can win the $(1:b)$ game on $\mathcal{F}$ for every $b > f_{\mathcal{F}}^+$. In the *monotone* game, there exists a unique threshold bias $f_{\mathcal{F}}^{\mathrm{mon}}$ for which Enforcer can win the $(1:b)$ game if and only if $b \leq f_{\mathcal{F}}^{\mathrm{mon}}$. Determining the order of magnitude of these threshold biases is a central problem in Avoider-Enforcer games. This appears to be a very difficult problem to solve in full generality, and a complete solution seems to be beyond our current means. Nevertheless, in the case of Avoider-Enforcer non-planarity game, we obtain essentially optimal bounds on the threshold biases, thus addressing a question and substantially improving the results of Hefetz, Krivelevich, Stojaković and Szabó [55]. The interested reader may wish to consult the book of Hefetz et al. [57] for recent progress in Avoider-Enforcer games, as well as its standing challenges and open problems.

The results of this chapter is joint work with Dennis Clemens, Julia Ehrenmüller and Yury Person [21].

Each subsequent chapter will contain its own introduction where backgrounds and motivations of the problems are discussed in more details. Any specific notation will be introduced in the individual chapters, but below we collect the standard notation used throughout this dissertation.

General notation

We use standard graph-theoretic notation and follow mainly the notation used in [13]. In particular, a graph is a pair $G = (V, E)$, where V is a finite set and $E \subseteq \binom{V}{2}$ is a subset of the pairs of elements of V. The elements in V are called *vertices*, and elements in E are called *edges*. Two vertices $v, w \in V$ are said to be *adjacent* if $\{v, w\} \in E$.

Let a graph G be given. Then we denote by $V(G)$ its set of vertices, and by $E(G)$ its set of edges. Their sizes are denoted with $v(G) = |V(G)|$ and $e(G) = |E(G)|$. Given a subset $U \subseteq V(G)$ of the vertices, we write $G[U]$ for the subgraph of G induced by the vertices of U. The *common neighbourhood* $N_G(U)$ of U is the set of all vertices of G that are adjacent to every vertex in U. For every vertex $v \in V(G)$ and every set $A \subseteq V(G)$, the set of all vertices in A adjacent to v is denoted by $N_G(v, A)$, and we write $\deg_G(v, A) = |N_G(v, A)|$ for the *degree of* v *in* A. The *degree of* v *in* G is $d_G(v) := \deg_G(v, V(G))$. By $\delta(G)$ we denote the minimum degree of G, the smallest degree a vertex in G can have. For two vertex sets $A, B \subseteq V(G)$, we let $e_G(A, B) = |\{(v, w) \in A \times B : \{v, w\} \in E(G)\}|$. The *edge density* between two disjoint sets $A, B \subset V(G)$ is denoted by $d_G(A, B) := \frac{e_G(A,B)}{|A||B|}$. Further standard graph parameters we shall use are the *independence number* $\alpha(G)$, the maximum size of a subset of the vertices without edges; the *chromatic number* $\chi(G)$, the smallest number k such that the vertices can be coloured with k colours so that no two vertices of the same colour are adjacent. Often, when the base graph G is clear from the context we omit the subscript G.

For $a, b, c \in \mathbb{R}$ we write $a = b \pm c$ if $b - c \leq a \leq b + c$. In order to simplify the presentation, we omit floors and ceilings and treat large numbers as integers whenever this does not affect the argument.

The set $\{1, 2, \dots, n\}$ of the first n positive integers is denoted by $[n]$. For $k \in \mathbb{N}$, we define $\binom{X}{k} := \{A \subseteq X : |A| = k\}$. We use the symbol $\dot{\bigcup}$ for union of disjoint sets.

We make use of asymptotic notation throughout the thesis. Given two functions $f, g : \mathbb{N} \to \mathbb{R}$, we write $f = O(g)$ if there is a constant $C > 0$ such that $f(n) \leq Cg(n)$ for all $n \in \mathbb{N}$. If $\lim_{n\to\infty} f(n)/g(n) = 0$, we write $f = o(g)$ and $g = \omega(f)$. Finally, unless stated otherwise all logarithms are to the base e.

Chapter 2

Removal and stability for Erdős-Ko-Rado

2.1 Introduction

In this chapter we derive a removal lemma for large families, showing that families of size close to $\ell\binom{n-1}{k-1}$ with relatively few disjoint pairs must be close to a union of ℓ stars. We then use this removal lemma to obtain a sparse version of the Erdős–Ko–Rado theorem.

We now discuss the Erdős–Ko–Rado Theorem and the history of these problems in greater detail, before presenting our new results.

2.1.1 Intersecting families and stability

A family $\mathcal{F} \subset \binom{[n]}{k}$ is said to be *intersecting* if $F_1 \cap F_2 \neq \emptyset$ for every $F_1, F_2 \in \mathcal{F}$. The natural extremal question is to ask how large such a family can be. When $n < 2k$, there are no two disjoint sets, and hence $\binom{[n]}{k}$ is intersecting. For $n \geq 2k$, a natural construction is to take all sets containing some fixed element $i \in [n]$. This family, called the *star with centre i*, contains $\binom{n-1}{k-1}$ sets, and Erdős, Ko and Rado [37] showed this is best possible.

Given the extremal result, great efforts have been made to better understand the general structure of large intersecting families. Hilton and Milner [59] determined the size of the largest intersecting family that is not a subset of a star, before Frankl [43] extended this to determine the size of the largest intersecting family not containing too large a star.

In the years since these initial papers appeared, a series of stability results have been obtained. Friedgut [47] and Dinur and Friedgut [27] used spectral techniques to show, provided

$k \le (\frac{1}{2} - \gamma)n$ for some $\gamma > 0$, any intersecting family of size close to $\binom{n-1}{k-1}$ is almost entirely contained in a star. Keevash and Mubayi [69] and Keevash [68] combined these methods with combinatorial arguments to provide similar results when k is close to $\frac{1}{2}n$.

However, a recent trend in extremal set theory is to go beyond the Erdős–Ko–Rado threshold and study set families that may not be intersecting, but contain few disjoint pairs. Das, Gan and Sudakov [23] studied the supersaturation problem, determining the minimum number of disjoint pairs appearing in sufficiently sparse k-uniform families. Furthermore, a probabilistic variant of this supersaturation problem was introduced by Katona, Katona and Katona [67], and further studied by Russell [89], Russell and Walters [90] and Das and Sudakov [24].

Another direction that has been pursued is the transferral of the Erdős–Ko–Rado theorem to the sparse random setting. This study was initiated by Balogh, Bohman and Mubayi [3], who asked when the largest intersecting subfamily of a random k-uniform hypergraph is the largest star. Progress on this problem has been made in subsequent papers by Gauy, Hàn and Oliveira [51], Balogh, Das, Delcourt, Liu and Sharifzadeh [5] and Hamm and Kahn [52, 53]. An alternative version of a sparse Erdős–Ko–Rado theorem, which we shall discuss in greater detail in Section 2.1.3, was introduced by Bollobás, Narayanan and Raigorodskii [15].

2.1.2 Removal lemmas for disjoint pairs

As these new problems go beyond the Erdős–Ko–Rado threshold, we require more robust forms of stability that apply not only to intersecting families, but also to families with few disjoint pairs. This motivated the search for a *removal lemma* that would show one can remove few sets from any family with a small number of disjoint pairs to obtain an intersecting family. Such a result would be the set-theoretic analogue of the graph removal lemmas that have found a wide range of applications in extremal graph theory, details of which are in the survey of Conlon and Fox [22].

Friedgut and Regev [49] proved the first such removal lemma, stated below.

Theorem 2.1.1 (Friedgut–Regev)**.** *Let $\gamma > 0$, and let k and n be positive integers satisfying $\gamma n \le k \le (\frac{1}{2} - \gamma)n$. Then for every $\varepsilon > 0$ there is a $\delta > 0$ such that any family $\mathcal{F} \subset \binom{[n]}{k}$ with at most $\delta |\mathcal{F}| \binom{n-k}{k}$ disjoint pairs can be made intersecting by removing at most $\varepsilon \binom{n-1}{k-1}$ sets from $\mathcal{F}$.*

This is a very general result that holds regardless of the size or structure of the nearest intersecting family. However, for extremal applications, one is typically interested in the case when $|\mathcal{F}| \approx \binom{n-1}{k-1}$. For example, Gauy, Hàn and Oliveira required such a lemma in [51], coupling Theorem 2.1.1 with known stability results to show that a family of size close to $\binom{n-1}{k-1}$ with few disjoint pairs must be close in structure to a star. They further asked if such a result also holds for $k = o(n)$. Our main theorem shows this is indeed the case. Theorem 2.1.2 provides a removal lemma that holds whenever $\mathcal{F}$ has size close to a union of ℓ full stars and has relatively few disjoint pairs. Moreover, when $\ell = 1$, this holds for the full range of $2 \leq k < \frac{n}{2}$.

Theorem 2.1.2 ([25]). *There is an absolute constant $C > 1$ such that if n, k and ℓ are positive integers satisfying $n > 2k\ell^2$, and $\mathcal{F} \subset \binom{[n]}{k}$ is a family of size $|\mathcal{F}| = (\ell - \alpha)\binom{n-1}{k-1}$ with at most $\left(\binom{\ell}{2} + \beta\right)\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs, where $\max(2\ell|\alpha|, |\beta|) \leq \frac{n-2k}{(20C)^2 n}$, then there is a family $\mathcal{S}$ that is the union of ℓ stars satisfying $|\mathcal{F}\Delta\mathcal{S}| \leq C((2\ell-1)\alpha + 2\beta)\frac{n}{n-2k}\binom{n-1}{k-1}$.*

Another feature of Theorem 2.1.2 is that, despite its simple proof, it provides quantitative control that is often sharp up to the constant. The distance from $\mathcal{F}$ to a union of ℓ stars is measured in terms of its size (parametrised by α), the number of disjoint pairs (parametrised by β), and how close k is to $\frac{1}{2}n$. When $\ell = 0$, taking $\beta = 0$ gives a stability result for intersecting families, and the bounds sharpen those given by Keevash and Mubayi [69] and Keevash [68].

For positive β, the bounds remain sharp up to the constant. If k is bounded away from $\frac{n}{2}$, then one may take a star and add $\alpha\binom{n-1}{k-1}$ sets from another star to obtain a family of size $(1+\alpha)\binom{n-1}{k-1}$ with $\alpha\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs that is $\alpha\binom{n-1}{k-1}$-far from a star. On the other hand, if $t = n - 2k = o(n)$, consider the anti-star $\binom{[n-1]}{k}$. This has size $\left(1 + \frac{t}{k}\right)\binom{n-1}{k-1}$, contains approximately $\frac{t}{n}\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs, and yet is approximately $\binom{n-1}{k-1}$-far from a star.

When $\ell \geq 2$, $\mathcal{F}$ is much larger than the Erdős–Ko–Rado bound, and hence we would expect $\mathcal{F}$ to contain many disjoint pairs. Das, Gan and Sudakov [23] have shown that, provided n is sufficiently large, a union of ℓ stars, which has approximately $\binom{\ell}{2}\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs, minimise the number of disjoint pairs in set families of this size. Theorem 2.1.2 provides stability for this supersaturation result, showing that families of comparable size with a similar number of disjoint pairs must be close in structure to a union of ℓ stars.

Finally, while we require $n > 2k$ when $\ell = 1$, we can do a bit better when ℓ is large: as ℓ tends to infinity, the bound on n can be lowered to $n > (\frac{1}{2} + o(1))k\ell^2$.

2.1.3 Erdős–Ko–Rado for sparse Kneser subgraphs

To demonstrate the usefulness of Theorem 2.1.2, we shall apply it to a problem of Bollobás, Narayanan and Raigorodskii [15] regarding an extension of the Erdős–Ko–Rado theorem to the sparse random setting. To define the problem at hand, we first need to introduce the Kneser graph and its connection to the Erdős–Ko–Rado theorem.

Given integers $1 \leq k \leq \frac{1}{2}n$, the Kneser graph $K(n, k)$ is defined on the vertex set $V = \binom{[n]}{k}$, with two k-sets $F, G \in \binom{[n]}{k}$ adjacent in $K(n, k)$ if and only if $F \cap G = \emptyset$. Since edges of $K(n, k)$ denote disjoint pairs in $\binom{[n]}{k}$, it follows that independent sets of $K(n, k)$ correspond directly to intersecting families in $\binom{[n]}{k}$. Thus the Erdős–Ko–Rado theorem, viewed from the perspective of the Kneser graph, shows $\alpha(K(n, k)) = \binom{n-1}{k-1}$ when $n \geq 2k$.

Bollobás, Narayanan and Raigorodskii [15] transferred the Erdős–Ko–Rado theorem to the random setting by considering not the entire Kneser graph $K(n, k)$, but rather random subgraphs of it. Given some probability $0 \leq p \leq 1$, let $K_p(n, k)$ denote the subgraph of $K(n, k)$ where every edge is retained independently with probability p. As $K_p(n, k) \subseteq K(n, k)$, we clearly have $\alpha(K_p(n, k)) \geq \alpha(K(n, k)) = \binom{n-1}{k-1}$. They then asked for which p we have equality.

In their paper, they showed the Erdős–Ko–Rado theorem is surprisingly robust when k is not too large with respect to n. In other words, we almost surely have $\alpha(K_p(n, k)) = \binom{n-1}{k-1}$ even for very small probabilities p (and thus very sparse subgraphs of $K(n, k)$). Furthermore, they exhibited a sharp threshold for when this sparse Erdős–Ko–Rado theorem holds.

Theorem 2.1.3 (Bollobás–Narayanan–Raigorodskii). *Fix $\varepsilon > 0$ and suppose $2 \leq k = o(n^{1/3})$. Let*

$$p_0 = \frac{(k+1)\log n - k \log k}{\binom{n-1}{k-1}}.$$

Then, as $n \to \infty$,

$$\mathbb{P}\left(\alpha(K_p(n, k)) = \binom{n-1}{k-1}\right) \to \begin{cases} 0 & \text{if } p \leq (1-\varepsilon)p_0 \\ 1 & \text{if } p \geq (1+\varepsilon)p_0 \end{cases}.$$

Moreover, for $p \geq (1+\varepsilon)p_0$, with high probability the largest independent sets are stars.

While observing that we may take $\varepsilon = O(k^{-1})$, they conjectured that the result should continue to hold provided $k = o(n)$. Partial progress was made by Balogh, Bollobás and Narayanan [4], who showed that for every $\gamma > 0$ there is some constant $c(\gamma) > 0$ such that if $k \leq (\frac{1}{2} - \gamma)n$ and $p \geq \binom{n-1}{k-1}^{-c(\gamma)}$, then $\alpha(K_p(n, k)) = \binom{n-1}{k-1}$ with high probability.

By applying Theorem 2.1.2, we obtain sharper results for large k, as given in the theorem below. For these larger values of k, it is convenient to present the critical probability in a different form to that of Theorem 2.1.3; note that p_c below is asymptotically equal to p_0 above when $k = o(n^{1/2})$.

Theorem 2.1.4 ([25]). *There is an absolute constant $C > 0$ such that the following holds. Let k and n be integers with $\omega(1) = k \le \frac{1}{2}(n-3)$, let $\varepsilon = \omega(k^{-1})$, and set*

$$p_c = \frac{\log\left(n\binom{n-1}{k}\right)}{\binom{n-k-1}{k-1}}.$$

Then, as $n \to \infty$, $\mathbb{P}\left(\alpha(K_p(n,k)) = \binom{n-1}{k-1}\right) \to 0$ if $p \le (1-\varepsilon)p_c$.

For $k \le \frac{n}{6C}$, if $p \ge (1+\varepsilon)p_c$, with high probability $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ and the stars are the only maximum independent sets. For $k \le \frac{1}{2}(n-3)$, the same conclusion holds for $p \ge \frac{2Cn}{n-2k}p_c$.

Theorem 2.1.4 exhibits a sharp threshold for $k \le \frac{n}{6C}$, thus extending Theorem 2.1.3 to k as large as linear in n. Furthermore, when $k \le (\frac{1}{2} - \gamma)n$, as considered in [4], $\frac{n}{n-2k} \le (2\gamma)^{-1}$, and so Theorem 2.1.4 determines the critical probability up to a constant factor. Finally, when k is close to $\frac{1}{2}n$, we find that the sparse version of the Erdős–Ko–Rado theorem still holds for very small edge probabilities; when $k = \frac{1}{2}(n-3)$, we almost surely have $\alpha(K_p(n,k)) = \binom{n-1}{k-1}$ even for $p = \Omega(n^{-1})$.

The remaining of this chapter is organised as follows. In Section 2.2 we prove our removal lemma, Theorem 2.1.2. We apply this result to the sparse Erdős–Ko–Rado problem in Section 2.3, where we prove Theorem 2.1.4. The final section contains some concluding remarks and open problems.

2.2 The removal lemma

In this section we prove our version of the removal lemma, Theorem 2.1.2. Our proof combines the work of Lovász [75] on the spectrum of the Kneser graph with an analytic result of Filmus [41] regarding approximations of Boolean functions on $\binom{[n]}{k}$. Before beginning with the proof, we shall introduce the necessary terminology.

Given a family of sets $\mathcal{F} \subset \binom{[n]}{k}$, the *characteristic function* $f : \binom{[n]}{k} \to \{0,1\}$ is a Boolean function indicating membership of the family, with $f(F) = 1$ if and only if $F \in \mathcal{F}$. We may embed $\binom{[n]}{k} \subset \{0,1\}^n$ into the n-dimensional hypercube, and thus think of f as being defined on the k-uniform *slice* of the cube $\{(x_1, \dots, x_n) \in \{0,1\}^n : \sum_i x_i = k\}$. A function f is *affine* if $f(x_1, x_2, \dots, x_n) = a_0 + \sum_{i \in [n]} a_i x_i$ for some constants a_i, $0 \le i \le n$. We will equip this space of functions with the L_2-norm with respect to the uniform measure on $\binom{[n]}{k}$, defining

$$\|f - g\|^2 = \mathbf{E}\left[|f-g|^2\right] = \frac{1}{\binom{n}{k}} \sum_{F \in \binom{[n]}{k}} |f(F) - g(F)|^2,$$

and say f and g are *ε-close* if $\|f-g\|^2 \leq \varepsilon$. Finally, to avail of the spectral results, which are traditionally phrased in terms of matrices and vectors, we shall abuse notation and identify a function $f : \binom{[n]}{k} \to \mathbb{R}$ with the vector $f = (f(F))_{F \in \binom{[n]}{k}}$ in $\mathbb{R}^{\binom{[n]}{k}}$. Note that the L_2-norm above arises from the standard inner product on $\mathbb{R}^{\binom{n}{k}}$.

The first step of our proof is the following lemma, which transfers the problem into the analytic framework set up above. The lemma shows that if a set family $\mathcal{F}$ is as in the statement of Theorem 2.1.2, then its characteristic function can be approximated well by an affine function.

Lemma 2.2.1. *Let n, k and ℓ be positive integers satisfying $n > 2k$, and let $\mathcal{F} \subset \binom{[n]}{k}$ be a family of size $|\mathcal{F}| = (\ell - \alpha)\binom{n-1}{k-1}$ with at most $\left(\binom{\ell}{2} + \beta\right)\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs. If $f : \binom{[n]}{k} \to \{0,1\}$ is the characteristic function of $\mathcal{F}$, then $\|f-g\|^2 \leq ((2\ell-1)\alpha + 2\beta)\frac{k}{n-2k}$ for some affine function $g : \binom{[n]}{k} \to \mathbb{R}$.*

To prove Lemma 2.2.1, we require some information on the spectrum of the Kneser graph. Let A denote the adjacency matrix of $K(n,k)$. In his celebrated paper on the Shannon capacity of graphs, Lovász [75, page 6] showed the eigenvalues of A are $\lambda_i = (-1)^i\binom{n-k-i}{k-i}$ for $0 \leq i \leq k$. Thus the largest eigenvalue is the degree of the vertices in the regular graph $K(n,k)$, $\lambda_0 = \binom{n-k}{k}$, while the smallest eigenvalue is $\lambda_1 = -\binom{n-k-1}{k-1}$. The second smallest eigenvalue is $\lambda_3 = -\binom{n-k-3}{k-3}$. Furthermore, the λ_0-eigenspace is one-dimensional, spanned by the constant function. The $(n-1)$-dimensional λ_1-eigenspace is spanned by the functions $x_i - \frac{k}{n}$. Hence the span of the λ_0- and λ_1-eigenspaces is precisely the space of affine functions. As A is a real symmetric matrix, its eigenspaces are orthogonal. Armed with these preliminaries, we can prove the lemma.

Proof of Lemma 2.2.1. Given the characteristic vector f of $\mathcal{F}$, write $f = f_0 + f_1 + f_2$, where f_0 and f_1 are the projections of f to the λ_0- and λ_1-eigenspaces respectively, and $f_2 = f - f_0 - f_1$. By the orthogonality of eigenspaces, we have $\|f\|^2 = \|f_0\|^2 + \|f_1\|^2 + \|f_2\|^2$. As f is a Boolean function, $\|f\|^2 = \mathbf{E}[f^2] = \mathbf{E}[f] = |\mathcal{F}| / \binom{n}{k} = (\ell - \alpha)\frac{k}{n}$. Thus, solving for $\|f_1\|^2$, we find $\|f_1\|^2 = (\ell-\alpha)\frac{k}{n} - \|f_0\|^2 - \|f_2\|^2$. Furthermore, since the λ_0-eigenspace is spanned by the constant function, $f_0 \equiv \mathbf{E}[f] = (\ell-\alpha)\frac{k}{n}$, and so $\|f_0\|^2 = \mathbf{E}\left[f_0^2\right] = (\ell-\alpha)^2\frac{k^2}{n^2}$.

As A is the adjacency matrix of the Kneser graph $K(n,k)$, and f is the characteristic function of the set family $\mathcal{F}$, it follows that $f^TAf = 2\mathrm{dp}(\mathcal{F})$. Using our bound on the number of disjoint pairs in $\mathcal{F}$,

$$\begin{aligned}\left(\ell^2 - \ell + 2\beta\right)\binom{n-1}{k-1}\binom{n-k-1}{k-1} &\geq 2\mathrm{dp}(\mathcal{F}) = f^TAf = f_0^TAf_0 + f_1^TAf_1 + f_2^TAf_2 \\ &\geq \lambda_0 f_0^T f_0 + \lambda_1 f_1^T f_1 + \lambda_3 f_2^T f_2.\end{aligned}$$

We divide through by $\binom{n}{k}$ to normalise, obtaining

$$\frac{\left(\ell^2-\ell+2\beta\right)k}{n}\binom{n-k-1}{k-1} \geq \binom{n-k}{k}\|f_0\|^2 - \binom{n-k-1}{k-1}\|f_1\|^2 - \binom{n-k-3}{k-3}\|f_2\|^2.$$

Dividing by $\binom{n-k-1}{k-1}$, substituting our expressions for $\|f_0\|^2$ and $\|f_1\|^2$, and simplifying gives

$$\begin{aligned}\frac{2\beta k}{n} &\geq \left[1-\frac{(k-1)(k-2)}{(n-k-1)(n-k-2)}\right]\|f_2\|^2 - \frac{(2\ell-1)\,\alpha k}{n} + \frac{\alpha^2 k}{n} \\ &= \frac{(n-2k)(n-3)}{(n-k-1)(n-k-2)}\|f_2\|^2 - \frac{(2\ell-1)\,\alpha k}{n} + \frac{\alpha^2 k}{n} \geq \frac{n-2k}{n}\|f_2\|^2 - \frac{(2\ell-1)\,\alpha k}{n}.\end{aligned}$$

Rearranging, we deduce $\|f_2\|^2 \leq ((2\ell-1)\alpha+2\beta)\frac{k}{n-2k}$. Recalling that f_0+f_1 is spanned by the λ_0- and λ_1-eigenspaces, and hence affine, setting $g = f_0+f_1$ gives the desired result. □

Lemma 2.2.1 shows the characteristic function of $\mathcal{F}$ must be close to an affine function, from which we shall deduce that $\mathcal{F}$ itself is close to a union of stars. Note that the characteristic function g of the union of stars with centres $i \in S$ is simply $g(x_1,\ldots,x_n) = \max_{i\in S} x_i$, and is thus determined only by the coordinates in S. The Friedgut–Kalai–Naor theorem [48] states that if a Boolean function $f : \{0,1\}^n \to \{0,1\}$ on the entire hypercube is close to an affine function, then it is close to a function determined by at most one coordinate. We shall make use of an analogous result for the k-uniform slices of the cube, due to Filmus [41, Theorem 3.1].

Theorem 2.2.2. *For some constant $C > 1$ the following holds. Suppose $2 \leq k \leq \frac{1}{2}n$ and $\varepsilon < \frac{k}{128n}$. For every Boolean function $f : \binom{[n]}{k} \to \{0,1\}$ that is ε-close to an affine function, there is some set $S \subset [n]$ of size $|S| \leq \max\left(1, \frac{Cn\sqrt{\varepsilon}}{k}\right)$ such that either f or $1-f$ is $(C\varepsilon)$-close to $\max_{i\in S} x_i$.*

We now have all the necessary ingredients to prove the removal lemma.

Proof of Theorem 2.1.2. Set $\varepsilon = ((2\ell-1)\alpha+2\beta)\frac{k}{n-2k}$, and take C as in Theorem 2.2.2. By our bounds on α and β, $\varepsilon < \frac{k}{128C^2n}$. If $\mathcal{F}$ is as in the statement of the theorem, then by Lemma 2.2.1 its characteristic function f is ε-close to an affine function. By Theorem 2.2.2, there is some $S \subset [n]$ such that f or $1-f$ is $(C\varepsilon)$-close to $\max_{i\in S} x_i$. Without loss of generality, we may assume $S = [s]$, where $s \leq \max\left(1, \frac{Cn\sqrt{\varepsilon}}{k}\right)$. Let $g_s = \max_{i\in[s]} x_i$, and let $\mathcal{G}_s = \binom{[n]}{k} \setminus \binom{[n]\setminus[s]}{k}$ be the family corresponding to this characteristic function.

Note that $\|f-g_s\|^2 = |\mathcal{F}\Delta\mathcal{G}_s| / \binom{n}{k}$, since for any set $F \in \binom{[n]}{k}$ we have

$$|f(F)-g_s(F)| = \begin{cases} 1 & \text{if } F \in \mathcal{F}\Delta\mathcal{G}_s \\ 0 & \text{otherwise} \end{cases}.$$

Hence we must have $|\mathcal{F}\Delta\mathcal{H}| \le C\varepsilon\binom{n}{k}$ for $\mathcal{H} = \mathcal{G}_s$ or $\mathcal{H} = \overline{\mathcal{G}_s}$, depending on whether it is f or $1-f$ that is $(C\varepsilon)$-close to g_s. There are six possibilities to consider:

(i) $\mathcal{H} = \mathcal{G}_s, s \le \ell - 1$	(ii) $\mathcal{H} = \mathcal{G}_s, s \ge \ell + 1$	(iii) $\mathcal{H} = \overline{\mathcal{G}_0}$
(iv) $\mathcal{H} = \overline{\mathcal{G}_s}, s \ge 2$	(v) $\mathcal{H} = \overline{\mathcal{G}_1}$	(vi) $\mathcal{H} = \mathcal{G}_\ell$

Since $\mathcal{G}_\ell$ is the union of ℓ stars, we wish to show that (vi) must hold. We first consider the sizes of $\mathcal{F}$ and $\mathcal{H}$ to eliminate cases (i)-(iv). Recall that $|\mathcal{F}| = (\ell - \alpha)\binom{n-1}{k-1}$, and, by our bound on α, $\ell - \alpha \in (\ell - \frac{1}{8}, \ell + \frac{1}{8})$. Since $||\mathcal{F}| - |\mathcal{H}|| \le |\mathcal{F}\Delta\mathcal{H}| \le C\varepsilon\binom{n}{k} < \frac{1}{8}\binom{n-1}{k-1}$, we must have $\left(\ell - \frac{1}{4}\right)\binom{n-1}{k-1} \le |\mathcal{H}| \le \left(\ell + \frac{1}{4}\right)\binom{n-1}{k-1}$.

We have $|\mathcal{G}_s| \le s\binom{n-1}{k-1}$, which is too small if $s \le \ell - 1$. On the other hand, observe that $\mathcal{G}_s$, the union of s stars, grows with s. Thus, when $s \ge \ell + 1$,

$$|\mathcal{G}_s| \ge |\mathcal{G}_{\ell+1}| \ge (\ell+1)\binom{n-1}{k-1} - \binom{\ell+1}{2}\binom{n-2}{k-2} \ge \left(\ell + 1 - \frac{\ell^2 k}{n}\right)\binom{n-1}{k-1} \ge \left(\ell + \frac{1}{2}\right)\binom{n-1}{k-1},$$

which is too large. This rules out cases (i) and (ii). We also have $\left|\overline{\mathcal{G}_0}\right| = \binom{n}{k} = \frac{n}{k}\binom{n-1}{k-1} \ge 2\ell^2\binom{n-1}{k-1}$, which is again too large, ruling out case (iii) as well.

To handle case (iv), we show that $\overline{\mathcal{G}_s}$ is too large when $s \ge 2$. Since $\left|\overline{\mathcal{G}_s}\right| = \binom{n-s}{k}$ is decreasing in s, it suffices to take $s = \frac{Cn\sqrt{\varepsilon}}{k}$. We indeed have too many sets, as

$$|\overline{\mathcal{G}_s}| = \binom{n-s}{k} \ge \left(1 - \frac{sk}{n}\right)\binom{n}{k} = \left(1 - C\sqrt{\varepsilon}\right)\frac{n}{k}\binom{n-1}{k-1} > \frac{3\ell^2}{2}\binom{n-1}{k-1} \ge \left(\ell + \frac{1}{2}\right)\binom{n-1}{k-1}.$$

The above argument does not immediately rule out case (v), since if $s = \max\left(1, \frac{Cn\sqrt{\varepsilon}}{k}\right) = 1$, we may not assume $s = \frac{Cn\sqrt{\varepsilon}}{k}$. However, the family $\overline{\mathcal{G}_1}$ is still too large when $\ell \ge 2$, as

$$|\overline{\mathcal{G}_1}| = \binom{n-1}{k} = \frac{n-k}{k}\binom{n-1}{k-1} \ge \left(2\ell^2 - 1\right)\binom{n-1}{k-1} > \left(\ell + \frac{1}{2}\right)\binom{n-1}{k-1}.$$

To rule out case (v) when $\ell = 1$, we consider the number of disjoint pairs in $\mathcal{F}$. Note that each of the $\binom{n-1}{k}$ sets in $\overline{\mathcal{G}_1}$ is disjoint from $\binom{n-k-1}{k}$ other sets in $\overline{\mathcal{G}_1}$, and hence $\mathrm{dp}(\overline{\mathcal{G}_1}) = \frac{1}{2}\binom{n-1}{k}\binom{n-k-1}{k}$. Moreover, removing t sets from $\overline{\mathcal{G}_1}$ can account for at most $t\binom{n-k-1}{k}$ disjoint pairs. If $\mathcal{F}$ were close to $\overline{\mathcal{G}_1}$, then $\left|\overline{\mathcal{G}_1} \setminus \mathcal{F}\right| \le C\varepsilon\binom{n}{k}$, and so

$$\mathrm{dp}(\mathcal{F}) \ge \mathrm{dp}(\mathcal{F} \cap \overline{\mathcal{G}_1}) \ge \left(\tfrac{1}{2} - \tfrac{C\varepsilon n}{n-k}\right)\binom{n-1}{k}\binom{n-k-1}{k} > \left(\tfrac{1}{2} - 2C\varepsilon\right)\binom{n-1}{k}\binom{n-k-1}{k}.$$

On the other hand, we assumed $\mathcal{F}$ has at most $\beta\binom{n-1}{k-1}\binom{n-k-1}{k-1}$ disjoint pairs, so we must have $\beta \ge \left(\frac{1}{2} - 2C\varepsilon\right)\frac{(n-k)(n-2k)}{k^2} > \frac{n-2k}{2n}$, contradicting our bound on β.

Thus we are only left with case (vi), where $\mathcal{H}$ is the union of ℓ stars $\mathcal{G}_\ell$, and, as required, we have $|\mathcal{F}\Delta\mathcal{G}_\ell| \le C\varepsilon\binom{n}{k} = C((2\ell-1)\alpha + 2\beta)\frac{n}{n-2k}\binom{n-1}{k-1}$. $\square$

2.3 Independence number of random Kneser subgraphs

In this section we prove Theorem 2.1.4, establishing an analogue of the Erdős–Ko–Rado theorem for sparse random subgraphs of the Kneser graph. We will show that below the threshold, there is with high probability an independent *superstar*: some star $\mathcal{S}$ and a set $F \notin \mathcal{S}$ such that $\mathcal{S} \cup \{F\}$ is independent in $K_p(n,k)$. The upper bound on the critical probability essentially follows from a union bound over all potential independent sets, where we shall be able to take advantage of the fine control afforded to us by Theorem 2.1.2 to obtain sharp results when k is large.

Proof of Theorem 2.1.4. First we establish the lower bound on the critical probability. Suppose $\varepsilon = \omega(k^{-1})$ and $p \leq (1-\varepsilon)p_c$. We wish to show that with high probability, stars can be extended to independent superstars in $K_p(n,k)$. Let $\mathcal{S}$ be the star with centre 1, and for every $1 \notin F \in \binom{[n]}{k}$ let $\mathcal{E}_F$ be the event that $\mathcal{S} \cup \{F\}$ is independent in $K_p(n,k)$.

Note that F is disjoint from $\binom{n-k-1}{k-1}$ sets in $\mathcal{S}$, and for $\mathcal{E}_F$ to hold none of these edges can appear in $K_p(n,k)$. Thus $\mathbb{P}(\mathcal{E}_F) = (1-p)^{\binom{n-k-1}{k-1}}$. Moreover, the events $\{\mathcal{E}_F : 1 \notin F\}$ depend on mutually disjoint sets of edges of $K(n,k)$, and are thus independent. Hence we can bound the probability that the stars are the largest independent sets of $K_p(n,k)$ by

$$\mathbb{P}\left(\alpha(K_p(n,k)) = \binom{n-1}{k-1}\right) \leq \mathbb{P}\left(\cap_F \overline{\mathcal{E}_F}\right) = \left(1-(1-p)^{\binom{n-k-1}{k-1}}\right)^{\binom{n-1}{k}}.$$

This bound is increasing in p, so it suffices to take $p = (1-\varepsilon)p_c = \frac{(1-\varepsilon)\log\left(n\binom{n-1}{k}\right)}{\binom{n-k-1}{k-1}}$. As $n \geq 2k+2$, $p = O(n^{-1}) = o(\varepsilon)$, and hence $(1-p)^{\binom{n-k-1}{k-1}} \geq \exp\left(-p(1+p)\binom{n-k-1}{k-1}\right) \geq \left(n\binom{n-1}{k}\right)^{-(1-\varepsilon/2)}$. Thus

$$\left(1-(1-p)^{\binom{n-k-1}{k-1}}\right)^{\binom{n-1}{k}} \leq \exp\left(-\binom{n-1}{k}(1-p)^{\binom{n-k-1}{k-1}}\right) \leq \exp\left(-n^{-1}\binom{n-1}{k}^{\varepsilon/2}\right) = o(1),$$

since $\varepsilon = \omega\left(k^{-1}\right)$. Hence for $p \leq (1-\varepsilon)p_c$ we have $\alpha(K_p(n,k)) > \binom{n-1}{k-1}$ with high probability.

We now seek an upper bound on the critical probability. By monotonicity, it suffices to consider p as small as possible. To begin, we shall prove the coarse threshold. Let C be the (absolute) constant from Theorem 2.1.2, and take $p = \zeta p_c$, where $\zeta = \frac{2Cn}{n-2k}$. For such p, we wish to show the only maximum independent sets of $K_p(n,k)$ are the stars. To this end, we define the following random variables:

$$X = |\{\text{independent superstars } \mathcal{F} : \mathcal{F} = \mathcal{S} \cup \{F\} \text{ for some star } \mathcal{S}, F \notin \mathcal{S}\}| \text{ and}$$
$$Y_i = \left|\left\{\text{independent } \mathcal{F} : |\mathcal{F}| = \binom{n-1}{k-1}, \min_{\mathcal{S} \text{ a star}} |\mathcal{S} \setminus \mathcal{F}| = i\right\}\right|,\ 1 \leq i \leq \binom{n-1}{k-1}.$$

X counts the number of independent superstars in $K_p(n,k)$. If $X=0$, the stars are all maximal independent sets. If we further have $Y_i=0$ for all $1\le i\le \binom{n-1}{k-1}$, then there are no independent sets of size $\binom{n-1}{k-1}$ that are not stars, and thus the stars are the only maximum independent sets in $K_p(n,k)$. Hence our task is to show $X+\sum_i Y_i=0$ with high probability. By the union bound, it suffices to show $\mathbb{P}(X>0)+\sum_i \mathbb{P}(Y_i>0)=o(1)$.

We begin by estimating $\mathbb{P}(X>0)$, which we can bound by $\mathbf{E}[X]$. There are n choices for the star $\mathcal{S}$, $\binom{n-1}{k}$ choices for the set $F\notin\mathcal{S}$, and, for each such configuration, $\binom{n-k-1}{k-1}$ edges that should not appear in $K_p(n,k)$, which occurs with probability $(1-p)^{\binom{n-k-1}{k-1}}\le \exp\left(-\zeta p_c\binom{n-k-1}{k-1}\right)$. Thus

$$\mathbf{E}[X]\le n\binom{n-1}{k}\exp\left(-\zeta p_c\binom{n-k-1}{k-1}\right)=\left(n\binom{n-1}{k}\right)^{1-\zeta}=o(1), \tag{2.1}$$

even for ζ as small as $1+\omega\left(k^{-1}\right)$.

To analyse $\mathbb{P}(Y_i>0)$, we shall distinguish between two different cases: families that are close to a star, and families far from a star. For the first case, we assume $1<i\le t_1=\frac{1}{400C}\binom{n-1}{k-1}$. The families $\mathcal{F}$ counted by Y_i have size $\binom{n-1}{k-1}$ and $|\mathcal{F}\Delta\mathcal{S}|=2\left|\mathcal{S}\setminus\mathcal{F}\right|\ge 2i$ for every star $\mathcal{S}$. By applying Theorem 2.1.2 with $\alpha=0$ and $\beta=\frac{i(n-2k)}{Cn\binom{n-1}{k-1}}$, it follows that $\mathrm{dp}(\mathcal{F})\ge\frac{i(n-2k)}{Cn}\binom{n-k-1}{k-1}$. For $\mathcal{F}$ to be independent in $K_p(n,k)$, none of these edges can appear, which occurs with probability $(1-p)^{\mathrm{dp}(\mathcal{F})}\le\left(n\binom{n-1}{k}\right)^{-\zeta i(n-2k)/(Cn)}$.

We now take a union bound over all possible choices of $\mathcal{F}$. We know there is some star $\mathcal{S}$ such that $|\mathcal{S}\setminus\mathcal{F}|=i$. There are n choices for the star $\mathcal{S}$, $\binom{\binom{n-1}{k-1}}{i}$ choices for the i sets in $\mathcal{S}\setminus\mathcal{F}$, and $\binom{\binom{n-1}{k}}{i}$ choices for the i sets in $\mathcal{F}\setminus\mathcal{S}$. Hence there are at most $n\binom{\binom{n-1}{k-1}}{i}\binom{\binom{n-1}{k}}{i}\le n\left(\frac{ke^2\binom{n-1}{k}^2}{(n-k)i^2}\right)^i$ families $\mathcal{F}$ that can be counted by Y_i. Thus we have

$$\sum_{i=1}^{t_1}\mathbb{P}(Y_i>0)\le\sum_{i=1}^{t_1}n\left(\frac{ke^2\binom{n-1}{k}^2}{(n-k)i^2}\left(n\binom{n-1}{k}\right)^{-\zeta(n-2k)/(Cn)}\right)^i\le\sum_{i=1}^{t_1}\frac{ne^{2i}}{(ni)^{2i}}=o(1). \tag{2.2}$$

where the second inequality follows from our choice of $\zeta=\frac{2Cn}{n-2k}$.

Finally, we bound $\mathbb{P}(Y_i>0)$ when $i>t_1$. Applying Theorem 2.1.2 with $\alpha=0$ and $\beta=\frac{n-2k}{(20C)^2n}$, any family $\mathcal{F}$ counted by $\sum_{i>t_1}Y_i$ must have $\mathrm{dp}(\mathcal{F})\ge\frac{n-2k}{(20C)^2n}\binom{n-1}{k-1}\binom{n-k-1}{k-1}$. Hence the probability of such an $\mathcal{F}$ being independent in $K_p(n,k)$ is $(1-p)^{\mathrm{dp}(\mathcal{F})}\le\left(n\binom{n-1}{k}\right)^{-\frac{\zeta(n-2k)}{(20C)^2n}\binom{n-1}{k-1}}$. Recalling our choice of $\zeta=\frac{2Cn}{n-2k}$, we apply a trivial union bound, summing over all $\binom{\binom{n}{k}}{\binom{n-1}{k-1}}\le$

$\left(\frac{ne}{k}\right)^{\binom{n-1}{k-1}}$ families of size $\binom{n-1}{k-1}$ to find, when $k \geq 600C$,

$$\sum_{i=t_1+1}^{\binom{n-1}{k-1}} \mathbb{P}(Y_i > 0) \leq \left(\frac{ne}{k} \left(n\binom{n-1}{k} \right)^{-1/(200C)} \right)^{\binom{n-1}{k-1}} \leq \left(\frac{ne}{k} \left(\frac{k}{n} \right)^{k/(200C)} \right)^{\binom{n-1}{k-1}} = o(1). \tag{2.3}$$

Combining (2.1), (2.2), and (2.3), we find that when $p \geq \zeta p_c$, $\mathbb{P}(X > 0) + \sum_i \mathbb{P}(Y_i > 0) = o(1)$, and so for such p, the maximum independent sets in $K_p(n,k)$ are precisely the stars.

We now prove the sharp threshold result, for which we must show that the same conclusion holds when $k \leq \frac{n}{6C}$ and $\zeta = 1 + \varepsilon$, for some small $\varepsilon = \omega(k^{-1})$. As previously stated, the bound from (2.1) holds with this smaller value of ζ. However, bounding $\sum_i \mathbb{P}(Y_i > 0)$ requires more careful analysis. We now split the sum into three parts: $1 \leq i \leq t_0 = \frac{\varepsilon}{2}\binom{n-k-1}{k-1}$, $t_0 + 1 \leq i \leq t_1 = \frac{1}{400C}\binom{n-1}{k-1}$, and $t_1 + 1 \leq i \leq \binom{n-1}{k-1}$.

For the latter two parts, we modify slightly our analysis of the above bounds. When a family is only moderately close to a star, with $t_0 + 1 \leq i \leq t_1$, we again use the bound in (2.2). We begin by observing that $i > t_0 = \frac{\varepsilon}{2}\binom{n-k-1}{k-1} = \frac{\varepsilon k}{2(n-k)}\binom{n-k}{k}$, and so

$$\frac{ke^2\binom{n-1}{k}^2}{(n-k)i^2} \leq \frac{4e^2(n-k)\binom{n-1}{k}^2}{\varepsilon^2 k \binom{n-k}{k}^2} \leq \frac{4e^2(n-k)}{\varepsilon^2 k}\left(1 + \frac{k}{n-2k}\right)^{2k} \leq \frac{4e^2(n-k)e^{k/(2C)}}{\varepsilon^2 k}.$$

Since $k \leq \frac{n}{6C}$, we may also bound $\left(n\binom{n-1}{k}\right)^{-\zeta(n-2k)/(Cn)} \leq \left(\frac{k}{n}\right)^{k/(2C)}$, and so the bases of the exponential summands in (2.2) are at most

$$\frac{4e^2(n-k)}{\varepsilon^2 k}\left(\frac{ek}{n}\right)^{k/(2C)} \leq \frac{4e^3}{\varepsilon^2 C^{k/(3C)}} < \frac{1}{2},$$

as we may assume $k \geq 6C$ and $\varepsilon \geq 13C^{-k/(6C)}$. This then implies $\sum_{i=t_0+1}^{t_1} \mathbb{P}(Y_i > 0) = o(1)$.

For families that are far from a star, we can re-estimate the upper bound in (2.3) to show

$$\sum_{i>t_1} \mathbb{P}(Y_i > 0) \leq \left(\frac{ne}{k}\left(n\binom{n-1}{k}\right)^{-\frac{\zeta(n-2k)}{(20C)^2 n}} \right)^{\binom{n-1}{k-1}} \leq \left(\frac{ne}{k}\left(\frac{k}{n}\right)^{\frac{k}{2(20C)^2}} \right)^{\binom{n-1}{k-1}} \leq \left(\frac{ek}{n}\right)^{\binom{n-1}{k-1}} = o(1),$$

assuming $k \geq (40C)^2$.

To complete the proof of the sharp threshold, we must demonstrate that we are unlikely to obtain independent families that are very close to stars, with $1 \leq i \leq t_0$. In this range, we repeat the analysis of Bollobás, Narayanan and Raigorodskii in [15], and instead consider *maximal* independent families in $K_p(n,k)$.

For $j \geq i \geq 0$, let $Z_{i,j}$ denote the number of maximal independent families $\mathcal{F}$ such that there is some star $\mathcal{S}$ with $|\mathcal{S} \setminus \mathcal{F}| = i$ and $|\mathcal{F} \setminus \mathcal{S}| = j$. Observe that if $\mathcal{F}$ is a family counted by the random variable Y_i, then for any maximal independent family $\mathcal{F}'$ containing $\mathcal{F}$, the family $\mathcal{F}'$ must be counted by $Z_{i',j}$ for some $i' \leq i$ and $j \geq i$. Furthermore, if $i' = 0$, this family $\mathcal{F}'$ contains superstars. This shows

$$\bigcup_{1 \leq i \leq t_0} \{Y_i > 0\} \subseteq \{X > 0\} \cup \bigcup_{\substack{1 \leq i \leq t_0 \\ j \geq i}} \{Z_{i,j} > 0\},$$

and so $\sum_{i=1}^{t_0} \mathbb{P}(Y_i > 0) \leq \mathbb{P}(X > 0) + \sum_{i=1}^{t_0} \sum_{j \geq i} \mathbb{P}(Z_{i.j} > 0)$. We already have $\mathbb{P}(X > 0) = o(1)$, and hence it suffices to show $\sum_{i=1}^{t_0} \sum_{j \geq i} \mathbf{E}[Z_{i,j}] = o(1)$.

Let $\mathcal{F}$ be a maximal independent family counted by $Z_{i,j}$. Let $\mathcal{S}$ be the corresponding star, $\mathcal{A} = \mathcal{S} \setminus \mathcal{F}$, and $\mathcal{B} = \mathcal{F} \setminus \mathcal{S}$. Thus we have $|\mathcal{A}| = i$ and $|\mathcal{B}| = j$. By virtue of $\mathcal{F}$ being independent, all of the edges between $\mathcal{B}$ and $\mathcal{S} \setminus \mathcal{A}$ must be missing in $K_p(n,k)$. As $\mathcal{F}$ is maximal, each $A \in \mathcal{A}$ must have an edge to some $B \in \mathcal{B}$, for otherwise $\mathcal{F} \cup \{A\}$ would be a larger independent family. In particular, this implies $\mathcal{B}$ is a subset of the union of the neighbourhoods in $K(n,k)$ of $A \in \mathcal{A}$.

There are thus n choices for the star $\mathcal{S}$, $\binom{\binom{n-1}{k-1}}{i}$ choices for $\mathcal{A}$, and at most $\binom{i\binom{n-k}{k}}{j}$ choices for $\mathcal{B}$. Each $A \in \mathcal{A}$ must retain at least one of its edges to $\mathcal{B}$, which occurs with probability at most jp, independently for each of the i sets. Furthermore, as every $B \in \mathcal{B}$ has $\binom{n-k-1}{k-1}$ neighbours in $\mathcal{S}$, there are at least $j\left(\binom{n-k-1}{k-1} - i\right)$ edges between $\mathcal{B}$ and $\mathcal{S} \setminus \mathcal{A}$ that must be missing. This gives

$$\mathbf{E}[Z_{i,j}] \leq n\binom{\binom{n-1}{k-1}}{i}\binom{i\binom{n-k}{k}}{j}(jp)^i (1-p)^{j\left(\binom{n-k-1}{k-1}-i\right)} = z_{i,j}.$$

We first observe that, for $i \leq t_0 = \frac{\varepsilon}{2}\binom{n-k-1}{k-1}$ and $j \geq i$,

$$\begin{aligned} \frac{z_{i,j+1}}{z_{i,j}} = \frac{\binom{i\binom{n-k}{k}}{j+1}(j+1)^i}{\binom{i\binom{n-k}{k}}{j}j^i}(1-p)^{\binom{n-k-1}{k-1}-i} &\leq \frac{\left(i\binom{n-k}{k} - j\right)e}{j+1}(1-p)^{\left(1-\frac{\varepsilon}{2}\right)\binom{n-k-1}{k-1}} \\ &\leq e\binom{n-k}{k}\left(n\binom{n-1}{k}\right)^{-\left(1+\frac{\varepsilon}{4}\right)} = o(1), \end{aligned}$$

since $p = \zeta p_c = \frac{(1+\varepsilon)\log\left(n\binom{n-1}{k}\right)}{\binom{n-k-1}{k-1}}$. Hence $\sum_{j\geq i} z_{i,j} \leq 2z_{i,i}$. Thus

$$\begin{aligned}
\sum_{i=1}^{t_0}\sum_{j\geq i}\mathbf{E}[Z_{i,j}] &\leq 2\sum_{i=1}^{t_0} z_{i,i} \\
&= 2n\sum_{i=1}^{t_0}\binom{\binom{n-1}{k-1}}{i}\binom{i\binom{n-k}{k}}{i}(ip)^i\,(1-p)^{i\left(\binom{n-k-1}{k-1}-i\right)} \\
&\leq 2n\sum_{i=1}^{t_0}\left(e^2\binom{n-1}{k-1}\binom{n-k}{k}p\,(1-p)^{\left(1-\frac{\varepsilon}{2}\right)\binom{n-k-1}{k-1}}\right)^i \\
&\leq 2n\sum_{t=1}^{t_0}\left(e^2\binom{n-1}{k-1}\binom{n-k}{k}\frac{(1+\varepsilon)\log\left(n\binom{n-1}{k}\right)}{\binom{n-k-1}{k-1}}\left(n\binom{n-1}{k}\right)^{-\left(1+\frac{\varepsilon}{4}\right)}\right)^i \\
&= 2n\sum_{i=1}^{t_0}\left(\frac{(1+\varepsilon)e^2\log\left(n\binom{n-1}{k}\right)}{n\left(n\binom{n-1}{k}\right)^{\frac{\varepsilon}{4}}}\right)^i = o(1),
\end{aligned}$$

completing the proof for the sharp threshold. □

2.4 Concluding remarks

In this chapter, we built on the work of Filmus [41] to develop a removal lemma for large set families with few disjoint pairs. We then used this to determine the threshold for random Kneser subgraphs having the Erdős–Ko–Rado property, thus answering a question of Bollobás, Narayanan and Raigorodskii [15].

Rather than the probabilistic problem considered above, one might instead ask the corresponding extremal question: how sparse can a spanning subgraph G of $K(n,k)$ be if $\alpha(G) = \binom{n-1}{k-1}$? A lower bound can be obtained by requiring the stars to be maximal independent sets. For every set $F \in \binom{[n]}{k}$, and every element $x \notin F$, there must be an edge between F and the star $\mathcal{S}_x$ centred at x, for otherwise $\mathcal{S}_x \cup \{F\}$ would be an independent set of size $\binom{n-1}{k-1}+1$. As each edge $\{F, F'\}$ covers k stars, it follows that F must have degree at least $\frac{n-k}{k}$, and hence $G \subseteq K(n,k)$ must have at least $\frac{n-k}{2k}\binom{n}{k}$ edges.

Perhaps surprisingly, this simple lower bound can be tight. If k divides n, then Baranyai's Theorem [6] gives a partition of $\binom{[n]}{k}$ into perfect matchings. In the Kneser graph, this corresponds to a partition of the vertices into cliques of size $\frac{n}{k}$. Let G be the subgraph consisting only of these cliques. Any independent set in G can contain at most one vertex from each clique, and hence $\alpha(G) \leq \frac{k}{n}\binom{n}{k} = \binom{n-1}{k-1}$. Furthermore, G is $\frac{n-k}{k}$-regular, matching the lower bound given previously.

Theorem 2.1.4 shows that for this bound on the independence number to hold in random graphs, they must be denser by a factor of at least $\log\left(n\binom{n-1}{k}\right)$. However, these random graphs have the additional property that the only maximum independent sets are the stars, which is not the case in the construction given above. One might be interested in the extremal problem with this stricter requirement, or in the case when k does not divide n.

Returning to the random setting, Devlin and Kahn [26] have recently established threshold results when $k \sim \frac{n}{2}$. It remains to exhibit a sharp threshold around p_c for $k > \frac{n}{6C}$. We believe that, perhaps for smaller k, a more precise hitting time result may hold. Consider the random process where one removes edges from the Kneser graph $K(n,k)$ one at a time, selecting at each step an edge uniformly at random from those that remain. Is it true that, with high probability, $\alpha(G) > \binom{n-1}{k-1}$ precisely when a superstar is born? The fact that the lower bound from the sharp threshold comes from these superstars suggests this might be the case.

More generally, given how central intersecting families are to extremal set theory, we believe the removal lemma should find many other applications. In a forthcoming paper with Balogh, Liu and Sharifzadeh, we obtain some supersaturation results using the removal lemma with $\ell \geq 1$, extending the results of [23]. We hope that the lemma might prove useful for other research directions as well.

Chapter 3

Erdős-Rothschild problem for intersecting families

3.1 Introduction

Given a k-uniform hypergraph F, the Turán number $\mathrm{ex}(n, F)$ of F is the maximum number of edges in a k-uniform hypergraph on n vertices that does not contain a copy of F. Determining these numbers is one of central problems in Extremal Combinatorics. Erdős and Rothschild [36] in 1974 proposed a novel twist to this problem: instead of considering hypergraphs with no copies of F, they were interested in edge-colourings (not necessarily proper) of hypergraphs with no monochromatic copies of F. They asked for $c_{r,F}(n)$ the maximum possible number of edge colourings of a hypergraph on n vertices with r colours without a monochromatic copy of F, and wondered whether this would lead to extremal configurations that are substantially different from those of the Turán problem. Note that as every edge colouring of any F-free hypergraph contains no monochromatic copies F, we see that $c_{r,F} \geq r^{\mathrm{ex}(n,F)}$ for all $r \geq 2$. In the case $r = 2$ and $F = K_s$, Erdős and Rothschild conjectured that the above estimate is tight for $n \geq n_0(s)$ sufficiently large. This conjecture was verified by Yuster [96] for $s = 3$. The full conjecture for all $s \geq 3$ was proved by Alon, Balogh, Keevash and Sudakov [2] who further showed that an analogous result holds for three colours. The author of [2] noted that when more than three colours are used, the behaviour of $c_{r,K_s}(n)$ changes, making its determination both harder and more interesting. Namely, it was shown in [2] that for $r \geq 4$ and $s \geq 3$, $c_{r,K_s}(n)$ is exponentially larger than $r^{\mathrm{ex}(n,K_s)}$. Pikhurko and Yilma [87] later provided some exact results for $r \geq 4$. Other authors (see [61, 62, 63, 64, 65, 71, 72, 73]) have address the Erdős-Rothschild problem in the cases of forbidden monochromatic matchings, stars, paths, trees and some other hypergraphs.

In this chapter, we study the Erdős-Rothschild problem for intersecting families of sets, vector spaces and permutations, extending the previous results in this direction. Given a family of sets, vector spaces or permutations, we define an (r,t)-*colouring* of the family to be an r-colouring of its members such that each colour class is t-intersecting[1]. The Erdős–Rothschild problem then asks which families maximise the number of (r,t)-colourings. The study of this problem was initiated by Hoppen, Kohayakawa and Lefmann [62]. Note that, in contrast to the triangle-free case, this problem is trivial when $r = 2$. Indeed, let $\mathcal{F}$ be any family, and let $\mathcal{F}' \subset \mathcal{F}$ be a maximal t-intersecting subfamily. For any $F \in \mathcal{F} \setminus \mathcal{F}'$, there must be some $F' \in \mathcal{F}'$ such that $\{F, F'\}$ is not t-intersecting. Thus in any $(2,t)$-colouring of $\mathcal{F}$, F and F' must receive opposite colours. It follows that every $(2,t)$-colouring of $\mathcal{F}$ is determined by its restriction to $\mathcal{F}'$, and hence there are at most $2^{|\mathcal{F}'|}$ $(2,t)$-colourings. On the other hand, any two-colouring of a t-intersecting family $\mathcal{G}$ is a $(2,t)$-colouring, giving precisely $2^{|\mathcal{G}|}$ such colourings. Hence the largest t-intersecting families also have the most $(2,t)$-colourings[2]. The problem is of interest, then, when $r \geq 3$.

In the following sections we shall review what are known before presenting our new results.

3.1.1 Permutations

Denote by S_n the symmetric group on $[n]$. A family of permutations $\mathcal{F} \subseteq S_n$ is called t-*intersecting* if any two permutations in $\mathcal{F}$ agree on at least t points; that is, for any $\sigma, \pi \in \mathcal{F}$, $|\{i \in [n] : \sigma(i) = \pi(i)\}| \geq t$. A natural example of a t-intersecting family $\mathcal{F} \subseteq S_n$ is a t-star, where there exist $i_1, \ldots, i_t \in [n]$ and $j_1, \ldots, j_t \in [n]$ such that for every $\sigma \in \mathcal{F}$, $\sigma(i_1) = j_1, \ldots, \sigma(i_t) = j_t$. Confirming a conjecture of Deza and Frankl [44], Ellis, Friedgut and Pipel [31] proved that, for n sufficiently large with respect to t, a t-intersecting family $\mathcal{F} \subseteq S_n$ has size at most $(n-t)!$, with equality only if $\mathcal{F}$ is a t-star.

Our first result is an Erdős-Rothschild-type extension of the aforementioned theorem of Ellis, Friedgut and Pipel.

Theorem 3.1.1 ([20])**.** *For every $t \geq 1$, there is an $n_0 = n_0(t)$ such that if $n \geq n_0$, then a family $\mathcal{F} \subseteq S_n$ can have at most $3^{(n-t)!}$ $(3,t)$-colourings, with equality if and only if $\mathcal{F}$ is a t-star.*

[1] A family $\mathcal{F}$ is called t-intersecting if and only the *intersection* of every pair in $\mathcal{F}$ has size at least t.
[2] With a little more work, one can often show uniqueness.

3.1.2 Set families

For $k \geq 2$ and $1 \leq t < k$, a k-uniform family $\mathcal{F}$ on the ground set $[n]$ is *t-intersecting* if every two sets of $\mathcal{F}$ share at least t elements. A family $\mathcal{F}$ is called a *t-star* if every sets in $\mathcal{F}$ contains a fixed set of t elements. The classic Erdős-Ko-Rado theorem [37] states that for n sufficiently large with respect to k and t, the largest t-intersecting k-uniform families on $[n]$ have $\binom{n-t}{k-t}$ edges. Frankl [42] and Wilson [95] later showed that $n \geq (t+1)(k-t+1)$ was the correct bound. Moreover, for $n > (t+1)(k-t+1)$ equality is attained only by t-stars.

We show that just beyond the bound $n \geq (t+1)(k-t+1)$, t-stars also maximise the number of $(3,t)$-colourings.

Theorem 3.1.2 ([20])**.** *Let $n, k \geq 64$ and $t \geq 1$ be integers such that $n \geq (t+1)(k-t+1)+\eta_{k,t}$, where*

$$\eta_{k,t} = \begin{cases} k + 12\log k & \text{for } t = 1, \\ 60\log k & \text{for } t = 2 \text{ and } k-t \geq 3, \\ 1 & \text{for } t \geq 3 \text{ and } k-t \geq 3, \\ 1531 & \text{for } t \geq 2 \text{ and } k-t = 2, \\ 1244k & \text{for } t \geq 2 \text{ and } k-t = 1. \end{cases}$$

Then, a k-uniform family $\mathcal{F}$ on $[n]$ can have at most $3^{\binom{n-t}{k-t}}$ $(3,t)$-colourings, with equality if and only if $\mathcal{F}$ is a t-star.

Observe that $\eta_{k,t} = 1$, which we have for most values of k and t, is the best possible result, as when $n = (t+1)(k-t+1)$ there exist maximum t-intersecting families which are not t-stars. However, there is no doubt that the case $t = 1$ is the most natural and interesting to study. Theorem 3.1.2 shows t-stars maximise the number of $(3,1)$-colourings when $n \geq 3k + 12\log k$.

We obtain the following result when colourings with four or more colours are considered.

Theorem 3.1.3 ([20])**.** *Let n, k and r be integers with $k \geq 2$, $r \geq 4$ and $n > Cr^2ke^{k^2/n}\log n$ for sufficiently large constant C. All k-uniform families on $[n]$ which maximise the number of $(r,1)$-colourings are unions of $\lceil r/3 \rceil$ 1-stars.*

Note that the same conclusion was obtained by Hoppen, Kohayakawa and Lefmann [62] for all k, r and all $n \geq n_0(k,r)$ sufficiently large[3]. While Theorem 3.1.3 allows k to be as large as $o(\sqrt{n\log n})$ when $r < \frac{n^{1/4}}{\log n}$. In the proof we utilise a stability version of the Erdős-Ko-Rado theorem due to Dinur and Friedgut [27]. This has prevented us from extending the present

[3]Their bounds seem to require $n = \Omega_r(k^3)$.

result from 1 to larger t as no analogue of the Dinur-Friedgut theorem has been established for $t > 1$ [4].

3.1.3 Vector spaces

Let $\mathbb{F}_q^n$ be an n-dimensional vector space over the finite field $\mathbb{F}_q$. A simple counting argument shows that the number of k-dimensional subspaces of $\mathbb{F}_q^n$ is given by the Gaussian binomial coefficient

$$\begin{bmatrix} n \\ k \end{bmatrix}_q := \prod_{i=0}^{k-1} \frac{q^{n-i}-1}{q^{k-i}-1}.$$

A family $\mathcal{F}$ of k-dimensional subspaces of $\mathbb{F}_q^n$ is called *t-intersecting* if $\dim(F_1 \cap F_2) \geq t$ for any two subspaces $F_1, F_2 \in \mathcal{F}$. Hsieh [66], and Frankl and Wilson [46] proved an Erdős-Ko-Rado type theorem for this setting, showing that for $n \geq 2k+1$, any t-intersecting family of k-dimensional subspaces of $\mathbb{F}_q^n$ has size at most $\begin{bmatrix} n-t \\ k-t \end{bmatrix}_q$. Moreover, the only families achieving equality are *t-stars*, consisting of all k-dimensional subspaces through a given t-dimensional subspace. For an alternate proof, see [18].

The results we obtain for permutations and hypergraphs can be extended to vector spaces as well, and here we prove 1-stars maximise the number of $(3,1)$-colourings.

Theorem 3.1.4 ([20]). *Suppose n, k and q are integers with $k \geq 7, n \geq 2k+1$ and $q \geq 2$. Then a family $\mathcal{V}$ of k-dimensional subspaces of $\mathbb{F}_q^n$ can have at most $3^{\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q}$ $(3,1)$-colourings, with equality if and only if $\mathcal{V}$ is a 1-star.*

The bound of $n \geq 2k+1$ for $q \geq 2$ and $k \geq 7$ in the theorem is tight as the 1-stars are not the unique extremal families for smaller values of n. We remark that for a fixed prime power q and integers $k > t \geq 1$, and sufficiently large n [5], Hoppen, Lefmann and Odermann [65] determined families of k-dimensional subspaces of $\mathbb{F}_q^n$ that maximise the number of $(3,t)$- and $(4,t)$-colourings.

3.2 Three-coloured families

In this section we prove our results for $(3,t)$-colourings of families. The first subsection is devoted to a general lemma, which gives a simple condition for the number of $(3,t)$-colourings

[4] Recently, we have managed to generalise Theorem 3.1.3 to cover the case $t > 1$. The proof will appear in a new version of [20].

[5] In contrast, our bound on n does not depend on q.

to be maximised by the largest t-intersecting families. In the subsequent subsections, we verify this condition in the settings of permutations, vector spaces and set families.

3.2.1 A general lemma

The following lemma, phrased in general terms that will be applicable in all of our settings, gives a simple condition for the number of $(3,t)$-colourings to be maximised by the largest t-intersecting families.

Lemma 3.2.1. *Let N_0 denote the size of the largest t-intersecting family, N_1 the size of the largest non-maximum t-intersecting family, and suppose two distinct maximum t-intersecting families can have at most N_2 members in common. Suppose further that there are at most M maximal t-intersecting families. Provided*

$$N_0 - \max(N_1, N_2) - 51 \log M > 0, \tag{3.1}$$

a family $\mathcal{F}$ can have at most 3^{N_0} $(3,t)$-colourings, with equality if and only if $\mathcal{F}$ is a maximum t-intersecting family.

Proof. First, for every t-intersecting family $\mathcal{I}$, fix an (arbitrary) assignment of a maximal t-intersecting family $\mathcal{M}(\mathcal{I})$ containing $\mathcal{I}$. Now let $\mathcal{F}$ be any family, and let $c(\mathcal{F})$ denote the number of $(3,t)$-colourings of $\mathcal{F}$. We wish to show $c(\mathcal{F}) \geq 3^{N_0}$ if only if $\mathcal{F}$ is itself a t-intersecting family of size N_0.

The colour classes of a $(3,t)$-colouring of $\mathcal{F}$ give rise to a partition $\mathcal{F} = \mathcal{I}_1 \sqcup \mathcal{I}_2 \sqcup \mathcal{I}_3$ into t-intersecting families. We can then map the $(3,t)$-colourings of $\mathcal{F}$ to triples of maximal intersecting families $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$, where $\mathcal{M}_i = \mathcal{M}(\mathcal{I}_i)$ for $1 \leq i \leq 3$. Let $c(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ denote the number of $(3,t)$-colourings of $\mathcal{F}$ mapped to the triple $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$.

Since there are at most M maximal t-intersecting families, by pigeonhole principle there exists a triple $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$ with $c(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3) \geq 3^{N_0} M^{-3}$. If $\mathcal{M}_1 = \mathcal{M}_2 = \mathcal{M}_3 = \mathcal{M}$ for some maximal intersecting family $\mathcal{M}$, then we have $\mathcal{F} \subset \mathcal{M}$, and so $c(\mathcal{F}) = 3^{|\mathcal{F}|} \leq 3^{|\mathcal{M}|} \leq 3^{N_0}$, with equality if and only if $\mathcal{F} = \mathcal{M}$ and $|\mathcal{M}| = N_0$. Hence we may assume $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$ are not the same.

We now seek to upper bound the number of $(3,t)$-colourings mapped to $(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3)$. Noting that a set $F \in \mathcal{F}$ can receive colour i only if $F \in \mathcal{M}_i$. Now let e_s denote the number of sets that are contained in exactly s families C_i's, where $1 \leq s \leq 3$. We then have $c(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3) \leq 1^{e_1} 2^{e_2} 3^{e_3} = 2^{e_2} 3^{e_3}$.

Since there are at least two distinct maximal families in $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$, we either have a non-maximum t-intersecting family or two distinct maximum t-intersecting families, and so $e_3 = |\mathcal{M}_1 \cap \mathcal{M}_2 \cap \mathcal{M}_3| \leq \max(N_1, N_2)$. We also have $2e_2 + 3e_3 \leq e_1 + 2e_2 + 3e_3 = |\mathcal{M}_1| + |\mathcal{M}_2| + |\mathcal{M}_3| \leq 3N_0$, and hence $e_2 \leq \frac{3}{2}(N_0 - e_3)$. Thus

$$c(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3) \leq 2^{\frac{3}{2}(N_0 - e_3)} 3^{e_3} = 2^{\frac{3}{2}N_0} \left(3 \cdot 2^{-\frac{3}{2}}\right)^{\max(N_1,N_2)}.$$

Since $c(\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3) \geq c(\mathcal{F})M^{-3}$, this gives

$$c(\mathcal{F}) \leq 2^{\frac{3}{2}N_0} \left(3 \cdot 2^{-\frac{3}{2}}\right)^{\max(N_1,N_2)} M^3 = 3^{N_0} \left(2^{\frac{3}{2}} \cdot 3^{-1}\right)^{N_0 - \max(N_1,N_2) - \frac{6 \log M}{2 \log 3 - 3 \log 2}} < 3^{N_0}$$

as $N_0 - \max(N_1, N_2) - 51 \log M > 0$. Therefore, the only families maximising the number of $(3,t)$-colourings are the maximum t-intersecting families. □

In order to obtain concrete results for permutations, vector spaces and set families, we must check that (3.1) holds. This will entail using an Erdős–Ko–Rado-type theorem to determine N_0, a Hilton–Milner-type theorem for N_1, and having appropriate bounds on N_2 and M. In the following subsections, we verify the inequality in each of these settings.

3.2.2 Permutations

In this section we shall prove Theorem 3.1.1. In light of Lemma 3.2.1, we need to verify condition (3.1).

Proof of Theorem 3.1.1. A theorem of Ellis, Friedgut and Pilpel [31, Theorem 3] shows that for n sufficiently large with respect to t, the largest t-intersecting subfamilies of S_n are t-stars. So we have $N_0 = (n-t)!$. Two distinct t-stars are either disjoint or fix at least $t+1$ elements, and so $N_2 = (n-t-1)!$. By the stability result of Ellis [29, Theorem 9], a non-maximum maximal t-intersecting family can contain at most $N_1 = (1 - 1/e + o(1))(n-t)!$ permutations. It follows that $\max(N_1, N_2) = (1 - 1/e + o(1))(n-t)!$. Finally, it was proved by Balogh, Das, Delcourt, Liu and Sharifzadeh [5, Proposition 3.1] that the number of maximal t-intersecting families of permutations is at most $M = n^{n2^{2n-2t+1}}$.

It remains to verify that (3.1) holds. We have, for large n,

$$N_0 - \max(N_1, N_2) - 51 \log M = (1/e + o(1))(n-t)! - 51n2^{2n-2t+1} \log n > 0,$$

since $(n-t)! \geq \left(\frac{n-t}{t}\right)^{n-t}$. Hence, by Lemma 3.2.1, a family $\mathcal{F} \subseteq S_n$ can have at most $3^{(n-t)!}$ $(3,t)$-colourings, with equality if and only if $\mathcal{F}$ is a t-star. □

3.2.3 Set families

We now turn our attention to set families, and seek to prove Theorem 3.1.2 via Lemma 3.2.1. The following inequality will be useful for our analysis:

$$\frac{\binom{a}{r}}{\binom{b}{r}} = \prod_{j=0}^{r-1} \frac{a-j}{b-j} \geq \left(\frac{a}{b}\right)^r \text{ for } a \geq b \geq r. \tag{3.2}$$

Proof of Theorem 3.1.2. We shall verify that, for n, k and t as in the statement of the theorem, the condition (3.1) holds.

We begin with the case $t = 1$. By the extremal result of Erdős, Ko and Rado [37, Theorem 1], the 1-stars are the largest intersecting families, and so $N_0 = \binom{n-1}{k-1}$. The intersection of two distinct 1-stars fixes 2 elements, and hence $N_2 = \binom{n-2}{k-2}$. The stability result of Hilton and Milner [59, Theorem 3] bounds the sizes of non-maximum maximal intersecting families by at most $N_1 = \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$. Hence $\max(N_1, N_2) \leq \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$. Finally, the number of maximal intersecting families can be bounded by $M = \binom{n}{k}^{\binom{2k-1}{k-1}}$, due to Balogh et al. [5, Proposition 4.1]. Since $\binom{n}{k} \leq 2^n$, we can estimate $\log M \leq n\binom{2k-1}{k-1}$. Therefore, using (3.2), we obtain

$$\begin{aligned} N_0 - \max(N_1, N_2) - 51 \log M &\geq \binom{n-k-1}{k-1} - 51n\binom{2k-1}{k-1} - 1 \\ &\geq \left(\left(\frac{n-k-1}{2k-1} \right)^{k-1} - 51n \right) \binom{2k-1}{k-1} - 1. \end{aligned}$$

For $t = 1$, we have $n \geq (t+1)(k-t+1) + \eta_{k,t} = 3k + 12 \log k$. It follows that

$$\left(\frac{n-k-1}{2k-1} \right)^{k-1} = \frac{n-k-1}{2k-1} \left(\frac{n-k-1}{2k-1} \right)^{k-2} \geq \frac{n}{4k} \left(1 + \frac{12 \log k}{2k-1} \right)^{k-2}$$

As $1 + x \geq \exp(6x/11)$ for $x \in [0, 1]$, we have

$$\left(1 + \frac{12 \log k}{2k-1} \right)^{k-2} \geq \exp\left(\frac{72(k-2) \log k}{22k} \right) \geq \exp(3 \log k) = k^3,$$

for $k \geq 64$. Thus, $N_0 - \max(N_1, N_2) - 51 \log M \geq \left(\frac{1}{4}nk^2 - 51\right) \binom{2k-1}{k-1} - 1 > 0$.

From now on we suppose $t \geq 2$. The theorem of Wilson [95] shows that for $n \geq (t+1)(k-t+1)$, the largest t-intersecting hypergraphs are t-stars, and so we have $N_0 = \binom{n-t}{k-t}$. The intersection of two distinct t-stars fixes at least $t+1$ elements, and so $N_2 = \binom{n-t-1}{k-t-1}$. By the stability theorem of Ahlswede and Khachatrian [1], a non-maximum maximal t-intersecting can have at most

$$N_1 = \begin{cases} \binom{n-t}{k-t} - \frac{n-(t+1)(k-t+1)}{n-t-1} \binom{n-t-1}{k-t} & \text{if } k \leq 2t+1, \\ \binom{n-t}{k-t} - \min\left(\frac{n-(t+1)(k-t+1)}{n-t-1} \binom{n-t-1}{k-t}, \binom{n-k-1}{k-t} - t \right) & \text{if } k \geq 2t+2 \end{cases}$$

elements. It follows that

$$\max(N_1, N_2) \le \binom{n-t}{k-t} - \frac{1}{n}\binom{n-k-1}{k-t} + n, \tag{3.3}$$

and for $k \le 2t+1$,

$$\max(N_1, N_2) \le \binom{n-t}{k-t} - \frac{n-(t+1)(k-t+1)}{n-t-1}\binom{n-t-1}{k-t}. \tag{3.4}$$

Finally, Balogh et al. [5, Proposition 4.1] bound the number of maximal intersecting families by $M = \binom{n}{k}^{\binom{2(k-t)+1}{k-t}}$.

We now handle the case $k-t=1$. Since $M = \binom{n}{k}^{\binom{2(k-t)+1}{k-t}} = \binom{n}{k}^3$, we can evaluate $\log M \le 3k\log\left(\frac{en}{k}\right)$. Combining this inequality with (3.4), one gets

$$\begin{aligned} N_0 - \max(N_1, N_2) - 51\log M &\ge \frac{n-(t+1)(k-t+1)}{n-t-1}\binom{n-t-1}{k-t} - 153k\log\left(\frac{en}{k}\right) \\ &= n - 2k - 153k\log\left(\frac{en}{k}\right). \end{aligned}$$

Since we are assuming $n \ge (t+1)(k-t+1) + \eta_{k,t} = 2k + \eta_{k,k-1} = 1246k$, this expression is increasing in n, and hence $N_0 - \max(N_1, N_2) - 51\log M \ge (1244 - 153\log(1246e))\, k > 0$.

We next treat the case when $t \ge 2$ and $k-t=2$. By using the inequalities $\log M \le \binom{2(k-t)+1}{k-t} n = 10n$, $n \ge (t+1)(k-t+1) + \eta_{k,k-2} > 3k$ and (3.4), we see that

$$N_0 - \max(N_1, N_2) - 51\log M \ge (\tfrac{1}{3}\eta_{k,k-2} - 510)n > 0.$$

It remains to verify (3.1) of Lemma 3.2.1 in the cases when $t \ge 2$ and $k-t \ge 3$. Using the trivial bound $\binom{n}{k} \le 2^n$, we obtain $\log M \le n\binom{2(k-t)+1}{k-t} \le 2n\binom{2(k-t)}{k-t}$. By (3.3) and noting that $n-k-1 \ge t(k-t) + \eta_{k,t}$, we have

$$\begin{aligned} N_0 - \max(N_1, N_2) - 51\log M &\ge \frac{1}{n}\binom{n-k-1}{k-t} - n - 102n\binom{2(k-t)}{k-t} \\ &\overset{(3.2)}{\ge} \left(\frac{1}{n}\left(\frac{n-k-1}{2(k-t)}\right)^{k-t} - 103n\right)\binom{2(k-t)}{k-t} \\ &\ge \left(\frac{n^{10}}{(4(k-t))^{11}}\left(\frac{t(k-t)+\eta_{k,t}}{2(k-t)}\right)^{k-t-11} - 103n\right)\binom{2(k-t)}{k-t}. \end{aligned} \tag{3.5}$$

If $t=2$, then $\eta_{k,t} = 60\log k$ and $\frac{t(k-t)+\eta_{k,t}}{2(k-t)} = 1 + \frac{30\log k}{k-2}$. Since $1+x \ge \exp(x/2)$ for $x \in [0, \frac{5}{2}]$, we have

$$(k-2)^{-11}\left(1 + \frac{30\log k}{k-2}\right)^{k-13} \ge k^{-11}\exp\left(\frac{15(k-13)\log k}{k-2}\right) \ge k^{-11}\exp(11\log k) = 1$$

as $k \geq 64$. Combining this inequality with (3.5) we find that

$$N_0 - \max(N_1, N_2) - 51 \log M \geq (4^{-11} n^{10} - 103n) \binom{2(k-t)}{k-t} > 0,$$

for $n \geq k \geq 64$.

If $t \geq 3$, then $(k-t)^{-11} \left(\frac{t(k-t)+\eta_{k,t}}{2(k-t)}\right)^{k-t-11} > (k-t)^{-11} \left(\frac{3}{2}\right)^{k-t-11}$. This expression attains the minimum at $k - t = 27$, and so from (3.5) it follows that

$$N_0 - \max(N_1, N_2) - 51 \log M \geq \left(108^{-11} 1.5^{16} n^{10} - 103n\right) \binom{2(k-t)}{k-t} > 0,$$

for $n \geq (t+1)(k-t+1) + 1 \geq 4k - 7 \geq 249$. □

3.2.4 Vector spaces

We conclude this section by proving Theorem 3.1.4.

Proof of Theorem 3.1.4. Before we begin our calculations, it will be useful to have some bounds on the Gaussian binomial coefficient. Observe that

$$q^{k(n-k)} \leq \begin{bmatrix} n \\ k \end{bmatrix}_q = \prod_{i=0}^{k-1} \frac{q^{n-i} - 1}{q^{k-i} - 1} \leq (2q)^{k(n-k)}. \tag{3.6}$$

By the extremal result of Hsieh [66, Theorem 4.4], the 1-stars are the largest intersecting families, and so $N_0 = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$. As the intersection of two distinct 1-stars fixes a 2-dimensional subspace, we must have $N_2 = \begin{bmatrix} n-2 \\ k-2 \end{bmatrix}_q$. On the other hand, by the stability result of Blokhuis, Brouwer, Chowdhury, Frankl, Mussche, Patkós and Szőnyi [11, Theorem 1.4], a maximal family that is not a star can have at most $N_1 = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix}_q + q^k$ subspaces, a quantity slightly larger than $\begin{bmatrix} n-2 \\ k-2 \end{bmatrix}_q$. Thus $\max(N_1, N_2) = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix}_q + q^k$. Finally, the work of Balogh et al. [5, Proposition 6.1] allows us to bound the number of maximal intersecting families by $M = \begin{bmatrix} n \\ k \end{bmatrix}_q^{\binom{2k-1}{k-1}}$. Putting these parameters together, we have

$$\begin{aligned}
N_0 - \max(N_1, N_2) - 51 \log M &\geq q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix}_q - q^k - 51 \binom{2k-1}{k-1} \log \begin{bmatrix} n \\ k \end{bmatrix}_q \\
&\overset{(3.6)}{\geq} q^{k(k-1)} \cdot q^{(k-1)(n-2k)} - q^k - 51 \cdot 4^k k(n-k) \log(2q) \\
&= q^{(k-1)(n-k)} - q^k - 51k(n-k)4^k \log(2q) \\
&\geq q^{(k-1)(n-k)} - q^k - 13n^2 4^k \log(2q).
\end{aligned} \tag{3.7}$$

Since $7 \leq k < n/2$ and $q \geq 2$, we may bound $q^{(k-1)(n-k)} \geq q^{3n}$, $q^k \leq q^{n/2}$, $4^k \leq q^{n-1}$ and $\log(2q) \leq q$, and hence the right hand side of (3.7) can be bounded from below by

$$q^{3n} - q^{n/2} - 13n^2 q^n > q^n - q^{n/2} > 0,$$

where in the first inequality we use the fact that $q^{2n} > 13n^2$ for all $q \geq 2$ and $n \geq 2k+1 \geq 15$. This completes the proof of Theorem 3.1.4. $\square$

Remark 3.2.2. We may apply the stability result of Ellis [30] to show that every family $\mathcal{V}$ of k-dimensional subspaces of $\mathbb{F}_q^n$ has at most $3^{\left[{n-t \atop k-t}\right]_q}$ $(3,t)$-colourings, with equality if and only if $\mathcal{V}$ is a t-star, provided $n \geq n_0(q,t,k)$ sufficiently large. This recovers a result of Hoppen, Lefmann and Odermann [65] mentioned in the introduction. By extending Lemma 3.2.1 from 3 to larger r, we can prove that, a family maximises the number of (r,t)-colourings only if it is an union of t-stars. This rough structure characterisation allows us to determine the optimal families for $r \geq 5$ and $t = 1$, and for $r \in \{5,6,9\}$ and $t > 1$. We refer the readers to [20] for the details.

3.3 Multicoloured families

In this section we shall investigate the number of (r,t)-colourings of set families for $r \geq 4$, obtaining a precise characterisation of the extremal families for many ranges of the parameters. We begin with an optimisation problem that motivates the constructions for the lower bound.

3.3.1 Optimisation problem

A maximum intersecting family has r^{N_0} $(r,1)$-colourings, since each of the N_0 members of the family can be receive any of the r colours. However, when $r \geq 4$, we can do better by distributing the colours between a larger number of maximum intersecting families. The following optimisation problem, earlier discussed in [62], suggests that it is optimal to take $\lceil r/3 \rceil$ maximum intersecting families, and assign three colours to as many of them as possible.

Lemma 3.3.1. *Let $r \geq 2$ be an integer. Consider the maximisation problem below, denoted by* $\mathrm{MAX}(r)$*,*

$$\begin{aligned} \underset{\vec{\alpha} \in \mathbb{N}^r}{\textit{maximise}} \quad & \mathrm{OBJ}(\vec{\alpha}) = \prod_{i:\alpha_i \neq 0} \alpha_i \\ \textit{subject to} \quad & \sum_{i=1}^{r} \alpha_i \leq r, \end{aligned}$$

and let $\mathrm{OPT}(r)$ *denote its optimal value. The following statements hold.*

(i) For a feasible vector $\vec{\alpha}$, either $\mathrm{OBJ}(\vec{\alpha}) = \mathrm{OPT}(r)$ *or* $\mathrm{OBJ}(\vec{\alpha}) \leq \frac{8}{9}\mathrm{OPT}(r)$*.*

(ii) Up to a permutation of coordinates, all optimal solutions take one of the following forms:

(a) $r \equiv 0 \pmod 3$*:* $\vec{\alpha} = (\underbrace{3, \dots, 3}_{r/3}, 0, \dots, 0)$*.*

(b) $r \equiv 1 \pmod 3$*:* $\vec{\alpha} = (\underbrace{2, 2, 3, \dots, 3}_{\lceil r/3 \rceil}, 0, \dots, 0)$ *or* $\vec{\alpha} = (\underbrace{4, 3, \dots, 3}_{\lfloor r/3 \rfloor}, 0, \dots, 0)$*.*

(c) $r \equiv 2 \pmod 3$*:* $\vec{\alpha} = (\underbrace{2, 3, \dots, 3}_{\lceil r/3 \rceil}, 0, \dots, 0)$*.*

Proof. We will prove the lemma by showing that, if $\mathrm{OBJ}(\vec{\alpha}) \geq \frac{8}{9}\mathrm{OPT}(r)$, then $\vec{\alpha}$ is as in (a), (b) or (c) of (ii).

Suppose $\alpha_i \geq 5$ for some i. Then, there exists $j \neq i$ such that $\alpha_j = 0$, as $\sum_{i=1}^r \alpha_i \leq r$. Consider a new vector $\vec{\alpha}'$, where we replace α_i and α_j with 3 and $\alpha_i - 3$ respectively. Clearly, $\vec{\alpha}'$ is feasible, and $\mathrm{OBJ}(\vec{\alpha}') = \frac{3(\alpha_i - 3)}{\alpha_i} \cdot \mathrm{OBJ}(\vec{\alpha}) \geq \frac{6}{5}\mathrm{OBJ}(\vec{\alpha}) > \mathrm{OPT}(r)$, a contradiction.

Hence we may assume every coordinate of $\vec{\alpha}$ is at most 4. Now suppose $\sum_{i=1}^r \alpha_i < r$. Since $\mathrm{OBJ}(\vec{\alpha}) \geq \frac{8}{9}\mathrm{OPT}(r) > 0$, we must have $\alpha_i \neq 0$ for some i. Let $\vec{\alpha}'$ be the vector formed by replacing α_i with $\alpha_i + 1$. Obviously, $\vec{\alpha}'$ is feasible, and $\mathrm{OBJ}(\vec{\alpha}') = \frac{\alpha_i+1}{\alpha_i}\mathrm{OBJ}(\vec{\alpha}) \geq \frac{5}{4}\mathrm{OBJ}(\vec{\alpha}) > \mathrm{OPT}(r)$, since $1 \leq \alpha_i \leq 4$ and $\mathrm{OBJ}(\vec{\alpha}) > \frac{8}{9}\mathrm{OPT}(r)$. This is a contradiction.

Suppose $\vec{\alpha}$ has two coordinates $\alpha_i = 1$ and $1 \leq \alpha_j \leq 4$. Form a new vector $\vec{\alpha}'$ by replacing these coordinates with 0 and $\alpha_j + 1$. Then, $\vec{\alpha}'$ is again feasible, and $\mathrm{OBJ}(\vec{\alpha}') = \frac{\alpha_j+1}{\alpha_j}\mathrm{OBJ}(\vec{\alpha}) \geq \frac{5}{4}\mathrm{OBJ}(\vec{\alpha}) > \mathrm{OPT}(r)$, contradicting our assumption.

Thus every coordinate must be either $0, 2, 3$ or 4. Replacing every 4 with two coordinates both equal to 2 preserves feasibility without changing its objective value. Suppose now we have at least three coordinates equal to 2. Form a new vector $\vec{\alpha}'$ by replacing those three coordinates with two coordinates equal to 3 and one coordinate equal to 0. The resulting vector $\vec{\alpha}'$ is still feasible, and $\mathrm{OBJ}(\vec{\alpha}') = \frac{9}{8}\mathrm{OBJ}(\vec{\alpha})$. Hence there can be at most two coordinates equal to 2, and, up to permutation of coordinates, there is only one option for every r.

This implies the optimal solutions have all non-zero coordinates equal to 3, except for perhaps one or two coordinates equal to 2, or one coordinate equal to 4, giving the characterisation in *(ii)*. $\square$

3.3.2 Extremal families

Given integers n, k, ℓ, we write $\mathcal{S}_{n,k,\ell} = \{F \in \binom{[n]}{k} : F \cap [\ell] \neq \emptyset\}$. In this section we shall show $\mathcal{S}_{n,k,\lceil r/3 \rceil}$ maximises the number of $(r,1)$-colourings, provided $n > Cr^2 k e^{k^2/n} \log n$. Below we write $c(\mathcal{F})$ for the number of $(r,1)$-colourings of the family $\mathcal{F}$. To show $\mathcal{S}_{n,k,\lceil r/3 \rceil}$ is an optimal family, we require a result of Dinur and Friedgut [27, Theorem 1.4].

Theorem 3.3.2. *There exists a constant $K > 0$ such that if $\mathcal{F} \subseteq \binom{[n]}{k}$ is an intersecting family then there exists $i \in [n]$ such that all but $K\binom{n-2}{k-2}$ of the sets in $\mathcal{F}$ contain i.*

We also need the following generalisation of the Erdős-Ko-Rado theorem (see [50] and [45, Theorem 3]).

Theorem 3.3.3. *Let n and k be integers such that $2 \leq k < n/2$. Let $\mathcal{A}$ and $\mathcal{B} \subseteq \binom{[n]}{k}$ be two families such that $A \cap B \neq \emptyset$ for every $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Then $\min(|\mathcal{A}|, |\mathcal{B}|) \leq \binom{n-1}{k-1}$.*

We also require the following version of Hoeffding's Inequality (see [60, Theorem 2]).

Theorem 3.3.4. *Let the random variables $X_1, \ldots, X_n$ be independent, with $0 \leq X_i \leq a$ for each i. Let $X = \sum_{i=1}^{n} X_i$. Then, for any $t > 0$,*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq 2\exp\left(-\frac{t^2}{a^2 n}\right).$$

In our notation, Theorem 3.1.3 gives $c(n,k,r,1) = c(\mathcal{S}_{n,k,\lceil r/3 \rceil})$. It will be useful to have a simple lower bound on $c(n,k,r,1)$. Indeed, let $(\alpha_1, \ldots, \alpha_{\lceil r/3 \rceil}, 0, \ldots, 0)$ be a vector which solves OPT(r). We can use α_i colours to colour the sets in $\{F \in \binom{[n]}{k} : \min F = i\}$ for every $1 \leq i \leq \lceil r/3 \rceil$. This gives rise to at least $(\prod_i \alpha_i)^{\binom{n-1}{k-1} - r\binom{n-2}{k-2}}$ $(r,1)$-colourings of $\mathcal{S}_{n,k,\lceil r/3 \rceil}$. Hence

$$c(n,k,r,1) \geq [\mathrm{OPT}(r)]^{\binom{n-1}{k-1} - r\binom{n-2}{k-2}}. \tag{3.8}$$

Proof of Theorem 3.1.3. Let $\mathcal{F}$ be an extremal system. As $c(\mathcal{F}) > 0$, by applying Theorem 3.3.2 we can assume without loss of generality that

$$|\{F \in \mathcal{F} : \min F > r\}| \leq Kr\binom{n-2}{k-2}. \tag{3.9}$$

Our proof splits into two main steps.

Step 1: Describe a typical colouring of $\mathcal{F}$.

Consider an $(r,1)$-colouring φ of $\mathcal{F}$. For each vertex $i \in [n]$ and each colour $\sigma \in [r]$, let $\mathcal{F}_{i,\sigma}$ denote the subfamily of $\mathcal{F}_i := \{F \in \mathcal{F} : i \in F\}$ induced by sets of colour σ. We say $\mathcal{F}_{i,\sigma}$ is *substantial* if $|\mathcal{F}_{i,\sigma}| \geq 3\binom{n-2}{k-2}$. The family $\mathcal{F}_i$ is called *s-influential* if there are precisely s colours $\sigma \in [r]$ for which $\mathcal{F}_{i,\sigma}$ is substantial.

Claim 3.3.5. *Suppose φ is an $(r,1)$-colouring of $\mathcal{F} \subset \binom{[n]}{k}$. Then, for each colour $\sigma \in [r]$ there is at most one vertex $i \in [n]$ such that $\mathcal{F}_{i,\sigma}$ is substantial.*

Proof. Assume to the contrary that $\mathcal{F}_{i,\sigma}$ and $\mathcal{F}_{j,\sigma}$ are substantial for some $i \neq j$. Let $\mathcal{A} = \{F : F \in \mathcal{F}_{i,\sigma}, j \notin F\}$ and $\mathcal{B} = \{F : F \in \mathcal{F}_{j,\sigma}, i \notin F\}$. Since φ is an $(r,1)$-colouring of $\mathcal{F}$,

$\mathcal{A}$ and $\mathcal{B}$ are cross-intersecting. It now follows from Lemma 3.3.3 that $\min(|\mathcal{A}|, |\mathcal{B}|) \le \binom{n-3}{k-2}$, and so $\min(|\mathcal{F}_{i,\sigma}|, |\mathcal{F}_{j,\sigma}|) \le \binom{n-3}{k-2} + \binom{n-2}{k-2} < 3\binom{n-2}{k-2}$. This contradicts our assumption that both $\mathcal{F}_{i,\sigma}$ and $\mathcal{F}_{j,\sigma}$ are substantial. $\square$

Given $r \in \mathbb{N}$, we denote by FEAS(r) the set of all vectors $\vec{\alpha} = (\alpha_1, \ldots, \alpha_r) \in \mathbb{N}^r$ with $\sum_{i=1}^r \alpha_i \le r$. For $\vec{\alpha} \in \mathbb{N}^r$ let $\mathcal{C}(\vec{\alpha})$ consist of all $(r,1)$-colourings φ of $\mathcal{F}$ such that $\mathcal{F}_i$ is α_i-influential with respect to φ for all $i \in [r]$. Then, Claim 3.3.5 implies

$$c(\mathcal{F}) = \sum_{\vec{\alpha} \in \text{FEAS}(r)} |\mathcal{C}(\vec{\alpha})| . \tag{3.10}$$

We shall show the contribution of $\mathcal{C}(\vec{\alpha})$ to $c(\mathcal{F})$ is negligible unless $\vec{\alpha}$ solves OPT(r). For those vectors $\vec{\alpha}$, we describe a typical colouring in $\mathcal{C}(\vec{\alpha})$. In order to do so we introduce a special class of colourings of $\mathcal{F}$.

Suppose $\vec{\alpha} \in$ OPT(r) and $\mathcal{P}(\vec{\alpha}) = \{(P_1, \ldots, P_r) : [r] = P_1 \dot\cup \ldots \dot\cup P_r \text{ and } |P_i| = \alpha_i \text{ for all } i\}$. For $P = (P_1, \ldots, P_r) \in \mathcal{P}(\vec{\alpha})$, we denote by $\mathcal{SC}(\vec{\alpha}, P)$ the set of all $(r,1)$-colourings φ of $\mathcal{F}$ with the following two properties

(P1) If $\varphi(e) \in P_i$, then $i \in e$;

(P2) If $\sigma \in P_i$, then σ appears at least $\frac{1}{4r}\binom{n-1}{k-1}$ times in $\mathcal{F}_i$.

The set of $\vec{\alpha}$-*star colourings* of $\mathcal{F}$ is defined as $\mathcal{SC}(\vec{\alpha}) = \bigcup_{P \in \mathcal{P}(\vec{\alpha})} \mathcal{SC}(\vec{\alpha}, P)$.

The relevance of star colourings is revealed in the next claim.

Claim 3.3.6. *Suppose* $|\mathcal{C}(\vec{\alpha})| \ge \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right) [\text{OPT}(r)]^{\binom{n-1}{k-1}}$, *then the following statements hold:*

(i) $\vec{\alpha}$ *solves* OPT(r),

(ii) $|\mathcal{F}_i| \ge \frac{1}{2}\binom{n-1}{k-1}$ *for all* $i \in \text{supp}(\vec{\alpha})$.

(iii) $\text{supp}(\vec{\alpha})$ *is a cover of* $\mathcal{F}$,

(iv) $|\mathcal{SC}(\vec{\alpha})| = (1 - o(1)) |\mathcal{C}(\vec{\alpha})|$.

Proof. By (3.10), it suffices to prove the statement for $\vec{\alpha} \in$ FEAS(r). Set $L := 3\binom{n-2}{k-2}$. Note that for each $i \in [r]$ the family $\mathcal{F}_i$ contains at most $\binom{n-1}{k-1}$ sets. The number of ways we may choose and colour some of them such that these selected and coloured sets are not substantial is at most

$$\sum_{(a_1,\ldots,a_r) \in [L]^r} \left(\prod_{s=1}^r \binom{|\mathcal{F}_i|}{a_s} \right) \le \left(\sum_{a \in [L]} \binom{\binom{n-1}{k-1}}{a} \right)^r \le \left(2\binom{\binom{n-1}{k-1}}{L} \right)^r \le \left(\frac{n}{k}\right)^{3r\binom{n-2}{k-2}}, \tag{3.11}$$

where the second inequality holds since $|\mathcal{F}_i| \le \binom{n-1}{k-1}$. Moreover, the number of ways we may colour sets in the family $\{F \in \mathcal{F} : \min F > r\}$ is bounded from above by

$$r^{|\{F\in\mathcal{F}:\min F>r\}|} \le r^{Kr\binom{n-2}{k-2}}, \tag{3.12}$$

as $|\{F \in \mathcal{F} : \min F > r\}| < Kr\binom{n-2}{k-2}$ by (3.9). On the other hand, the $(\alpha_1 + \ldots + \alpha_r)$ colours that make the families $\mathcal{F}_1, \ldots, \mathcal{F}_r$ influential can be selected and distributed among the families $\mathcal{F}_1, \ldots, \mathcal{F}_r$ in at most

$$\binom{r}{\alpha_1 + \ldots + \alpha_r} \frac{(\alpha_1 + \ldots + \alpha_r)!}{\alpha_1! \cdots \alpha_r!} < r^r \tag{3.13}$$

ways. To estimate $|\mathcal{C}(\vec{\alpha})|$ it remains to bound the number of ways the remaining sets in $\mathcal{F}_i$ may be coloured with any of the α_i colours that makes $\mathcal{F}_i$ α_i-influential. Trivially, the remaining sets in $\mathcal{F}_i$ can be coloured in at most

$$\max\{\alpha_i^{|\mathcal{F}_i|}, 1\} \le \max\left\{\alpha_i^{\binom{n-1}{k-1}}, 1\right\} \tag{3.14}$$

ways. In most of the cases this bound will suffice for our purpose. Now our analysis is divided into four cases.

(i) Assume to a contrary that $\vec{\alpha}$ does not solve OPT(r). Combining (3.11), (3.12), (3.13) and (3.14), we find

$$\begin{aligned}|\mathcal{C}(\vec{\alpha})| &\le \exp\left(3r^2\binom{n-2}{k-2}\log\left(\frac{n}{k}\right) + K\binom{n-2}{k-2}r\log r + r\log r\right)[\mathrm{OBJ}(\vec{\alpha})]^{\binom{n-1}{k-1}} \\ &\le \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}},\end{aligned}$$

as $\mathrm{OBJ}(\vec{\alpha}) \le \frac{8}{9}\mathrm{OPT}(r)$ by Lemma 3.3.1, and $n > Cr^2ke^{k^2/n}\log n$ by the hypothesis. This is a contradiction.

(ii) Suppose otherwise that $|\mathcal{F}_{i_0}| \le \frac{1}{2}\binom{n-1}{k-1}$ for some $i_0 \in \mathrm{supp}(\vec{\alpha})$. We can infer from (i) that $\vec{\alpha}$ solves OPT(r), and hence $\alpha_{i_0} \ge 2$ by Lemma 3.3.1. It follows that the number of ways the sets in $\mathcal{F}_{i_0}$ may be coloured with any of the α_{i_0} colours chosen in (3.13) is at most

$$\alpha_{i_0}^{|\mathcal{F}_{i_0}|} \le 2^{-\frac{1}{2}\binom{n-1}{k-1}} \cdot \alpha_{i_0}^{\binom{n-1}{k-1}}. \tag{3.15}$$

Combining (3.11), (3.12), (3.13), (3.14) and (3.15), we conclude

$$\begin{aligned}|\mathcal{C}(\vec{\alpha})| &\le \exp\left(3r^2\binom{n-2}{k-2}\log\left(\frac{n}{k}\right) + K\binom{n-2}{k-2}r\log r + r\log r - \frac{\log 2}{2}\cdot\binom{n-1}{k-1}\right) \\ &\quad \times [\mathrm{OBJ}(\vec{\alpha})]^{\binom{n-1}{k-1}} \\ &\le \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}},\end{aligned}$$

when $n > Cr^2ke^{k^2/n}\log n$. This contradicts our assumption.

(iii) Suppose for a contradiction that $\mathrm{supp}(\vec{\alpha})$ is not a cover of $\mathcal{F}$. Again by (i) we conclude $\vec{\alpha}$ solves OPT(r). Since $\mathrm{supp}(\vec{\alpha})$ is not a cover of $\mathcal{F}$, there exists a set $F \in \mathcal{F}$ with $F \cap \mathrm{supp}(\vec{\alpha}) = \emptyset$. Suppose F receives colour σ and $i_0 \in \mathrm{supp}(\vec{\alpha})$ is the vertex for which $\mathcal{F}_{i_0,\sigma}$ is substantial. Such a vertex i_0 exists since $\mathcal{F}$ is $\vec{\alpha}$-influential[6] and $\sum_i \alpha_i = r$. We shall bound the number of ways we can colour the sets in $\mathcal{F}_{i_0}$ with α_{i_0} colours selected in (3.13). Indeed, there are at most $\binom{n-k-1}{k-1}$ sets in $\mathcal{F}_{i_0}$ which are disjoint from F. Moreover, we can not use the colour of F for those sets. Thus, we can colour them in at most $(\alpha_{i_0}-1)^{\binom{n-k-1}{k-1}}$ ways. On the other hand, there are at most $\binom{n-1}{k-1} - \binom{n-k-1}{k-1}$ sets in $\mathcal{F}_{i_0}$ that intersect F. These sets can be coloured in at most $\alpha_{i_0}^{\binom{n-1}{k-1}-\binom{n-k-1}{k-1}}$ ways. Therefore, the number of ways we can colour the sets in $\mathcal{F}_{i_0}$ is bounded from above by

$$(\alpha_{i_0}-1)^{\binom{n-k-1}{k-1}} \cdot \alpha_{i_0}^{\binom{n-1}{k-1}-\binom{n-k-1}{k-1}} \leq \left(\tfrac{3}{4}\right)^{\binom{n-k-1}{k-1}} \alpha_{i_0}^{\binom{n-1}{k-1}}, \tag{3.16}$$

since $\alpha_{i_0} \leq 4$ by Lemma 3.3.1. Finally, it follows from (3.11), (3.12), (3.13), (3.14) and (3.16) that

$$\begin{aligned}|\mathcal{C}(\vec{\alpha})| &\leq \exp\left(3r^2\binom{n-2}{k-2}\log\binom{n}{k} + K\binom{n-2}{k-2}r\log r + r\log r + \log\left(\tfrac{3}{4}\right)\cdot\binom{n-k-1}{k-1}\right) \\ &\quad\times [\mathrm{OBJ}(\vec{\alpha})]^{\binom{n-1}{k-1}} \\ &\leq \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}},\end{aligned}$$

for $n > Cr^2ke^{k^2/n}\log n$. This contradicts the assumption.

(iv) The assertion can be restated as $\left|\mathcal{C}(\vec{\alpha}) \setminus \bigcup_{P\in\mathcal{P}(\vec{\alpha})}\mathcal{SC}(\vec{\alpha},P)\right| = o(|\mathcal{C}(\vec{\alpha})|)$. By the definition, $\mathcal{C}(\vec{\alpha}) \setminus \mathcal{SC}(\vec{\alpha},P)$ consists of all colourings $\varepsilon \in \mathcal{C}(\vec{\alpha})$ satisfying one of the following properties

($\overline{P1}$) There exist $i \in [r]$, $\sigma \in P_i$ and $e \in \mathcal{F}\setminus\mathcal{F}_i$ with $\varphi(e) = \sigma$;

($\overline{P2}$) There are $i \in [r]$ and $\sigma \in P_i$ such that σ appears less than $\frac{1}{4r}\binom{n-1}{k-1}$ times in $\mathcal{F}_i$.

To bound the number of colourings which satisfy $\overline{P1}$, we use the same argument as in (iii). The number of colourings with property $\overline{P2}$ can be bounded from above via Hoeffding's inequality. □

For a vector $\vec{\alpha} \in \mathbb{N}^r$ which solves OPT(r), we write $[\vec{\alpha}]$ for the set of all vectors $\vec{\beta} \in \mathbb{N}^r$ such that $\vec{\beta}$ solves OPT(r) and $\mathrm{supp}(\vec{\beta}) = \mathrm{supp}(\vec{\alpha})$. A typical colouring of $\mathcal{F}$ is described in the following statement.

[6]In other words, $\mathcal{F}_i$ is α_i-influential for every $i \in [r]$.

Claim 3.3.7. *There exists a vector $\vec{\alpha} \in \mathbb{N}^r$ such that*

$$|\mathcal{C}(\vec{\alpha})| \geq \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}} \text{ and } \sum_{\vec{\beta}\in[\vec{\alpha}]}\left|\mathcal{SC}(\vec{\beta})\right| = (1-o(1))c(\mathcal{F}).$$

Proof. Recall that $\mathrm{FEAS}(r) = \{(\alpha_1, \ldots, \alpha_r) \in \mathbb{N}^r \,|\, \sum_{i=1}^r \alpha_i \leq r\}$. Hence

$$|\mathrm{FEAS}(r)| = \binom{2r}{r} < 4^r. \tag{3.17}$$

We infer from (3.8), (3.10) and (3.17) that there exists a vector $\vec{\alpha} \in \mathrm{FEAS}(r)$ with

$$|\mathcal{C}(\vec{\alpha})| \geq 4^{-r} \cdot [\mathrm{OPT}(r)]^{\binom{n-1}{k-1} - r\binom{n-2}{k-2}} > \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}}. \tag{3.18}$$

Let $V = \left\{\vec{\beta} \in \mathbb{N}^r : \left|\mathcal{C}(\vec{\beta})\right| \geq \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}}\right\}$. It follows from (3.18) and Claim 3.3.6 that $V \subseteq [\vec{\alpha}]$, and so

$$\sum_{\vec{\beta}\in[\vec{\alpha}]}\left|\mathcal{SC}(\vec{\beta})\right| \geq \sum_{\vec{\beta}\in V}\left|\mathcal{SC}(\vec{\beta})\right| = (1-o(1))\sum_{\vec{\beta}\in V}\left|\mathcal{C}(\vec{\beta})\right| \tag{3.19}$$

by Claim 3.3.6 (iv). On the other hand,

$$\sum_{\vec{\beta}\notin V}\left|\mathcal{C}(\vec{\beta})\right| \overset{(3.17)}{\leq} 4^r \exp\left(-\frac{1}{4}\binom{n-k-1}{k-1}\right)[\mathrm{OPT}(r)]^{\binom{n-1}{k-1}} \overset{(3.8)}{=} o(c(\mathcal{F})) \tag{3.20}$$

for $n > Cr^2ke^{k^2/n}\log n$. Finally, by combining (3.10), (3.19), (3.20), we obtain

$$\sum_{\vec{\beta}\in[\vec{\alpha}]}\left|\mathcal{SC}(\vec{\beta})\right| = (1-o(1))c(\mathcal{F}),$$

as desired. □

Step 2: Show that $\mathcal{S}_{n,k,\lceil r/3\rceil}$ is optimal.

Claim 3.3.8. *Let $\vec{\alpha}$ be the vector as in Claim 3.3.7. Then $\mathcal{F} = \{F \in \binom{[n]}{k} : F \cap \mathrm{supp}(\vec{\alpha}) \neq \emptyset\}$.*

Proof. For simplicity of notation, let $\mathcal{S} = \{F \in \binom{[n]}{k} : F \cap \mathrm{supp}(\vec{\alpha}) \neq \emptyset\}$. Then, it follows from (3.18) and Claim 3.3.6 that $\mathcal{F} \subset \mathcal{S}$. Thus, $\mathcal{F} \neq \mathcal{S}$ if and only if there is $F \in \mathcal{S}_i \setminus \mathcal{F}$ for some $i \in \mathrm{supp}(\vec{\alpha})$. Assume this is the case. To get a contradiction we prove $c(\mathcal{F} \cup \{F\}) \geq (2-o(1))c(\mathcal{F})$. Note that, by Claim 3.3.7, we must have $\sum_{\vec{\beta}\in[\vec{\alpha}]}\left|\mathcal{SC}(\vec{\beta})\right| = (1-o(1))c(\mathcal{F})$. So it is enough to construct a 1-to-2 map between $\bigcup_{\vec{\beta}\in[\vec{\alpha}]}\mathcal{SC}(\vec{\beta})$ and the set of $(r,1)$-colourings of $\mathcal{F} \cup \{F\}$. Indeed, let φ be a colouring in $\bigcup_{\vec{\beta}\in[\vec{\alpha}]}\mathcal{SC}(\vec{\beta})$, say $\varphi \in \mathcal{SC}(\vec{\beta}, P)$ for some $\vec{\beta} \in [\vec{\alpha}]$ and $P = (P_1, \ldots, P_r) \in \mathcal{P}(\vec{\beta})$. Notice that

$$\left|\mathcal{F}_i \cap \bigcup_{j\in\mathrm{supp}(\vec{\alpha})\setminus\{i\}}\mathcal{F}_j\right| \leq r\binom{n-2}{k-2}. \tag{3.21}$$

On the other hand, since $\vec{\beta}$ solves OPT(r) and since $\varphi \in \mathcal{SC}(\vec{\beta}, P)$, we can infer from Lemma 3.3.1 and the definition of $\mathcal{SC}(\vec{\beta}, P)$ that $|P_i| \geq 2$ and each colour in P_i appears at least $\frac{1}{4r}\binom{n-1}{k-1} > r\binom{n-2}{k-2}$ times in $\mathcal{F}_i$. Combining this with (3.21), we conclude that at least two colours, say σ and τ, appear in $\mathcal{F}_i \setminus \bigcup_{j\in \mathrm{supp}(\vec{\alpha})\setminus\{i\}} \mathcal{F}_j$. Since $\varphi \in \mathcal{SC}(\vec{\beta})$, all the elements assigned either colour σ or colour τ by φ must contain i, and hence φ can be extended to two $(r,1)$-colourings of $\mathcal{F} \cup \{F\}$ by assigning either σ or τ to F. This finishes our proof of the claim. $\square$

Claim 3.3.9. $\mathcal{F} \cong \mathcal{S}_{n,k,\lceil r/3\rceil}$.

Proof. By Claim 3.3.8 and Lemma 3.3.1, we must have either $\mathcal{F} \cong \mathcal{S}_{n,k,\lceil r/3\rceil}$ or $\mathcal{F} \cong \mathcal{S}_{n,k,\lceil r/3\rceil-1}$. Assume to a contrary that $\mathcal{S}_{n,k,\lceil r/3\rceil-1}$ maximises the number of $(r,1)$-colourings. From this it follows from Claims 3.3.6 – 3.3.8 and Lemma 3.3.1 that $r = 3s+1$ for some integer $s \geq 1$. We then can deduce from Lemma 3.3.1 and Claim 3.3.7 that

$$c(\mathcal{S}_{n,k,\lceil r/3\rceil-1}) = (1+o(1))s \cdot \frac{(3s+1)!}{24\cdot 3^{s-1}} \cdot \left(7^{s-1}\cdot 6^{\binom{s-1}{2}}\right)^{(1+o(1))\binom{n-2}{k-2}} \cdot \left(4\cdot 3^{s-1}\right)^{\binom{n-1}{k-1}-(s-1)\binom{n-2}{k-2}},$$

and

$$c(\mathcal{S}_{n,k,\lceil r/3\rceil}) = (1+o(1))\binom{s+1}{2}\frac{(3s+1)!}{4\cdot 3^{s-1}}\cdot\left(4\cdot 5^{2(s-1)}\cdot 6^{\binom{s-1}{2}}\right)^{(1+o(1))\binom{n-2}{k-2}}\cdot\left(4\cdot 3^{s-1}\right)^{\binom{n-1}{k-1}-s\binom{n-2}{k-2}},$$

for $n > Cr^2ke^{k^2/n}\log n$. So $c(\mathcal{S}_{n,k,\lceil r/3\rceil}) > c(\mathcal{S}_{n,k,\lceil r/3\rceil-1})$, which contradicts our assumption. $\square$

Claim 3.3.9 shows that $\mathcal{S}_{n,k,\lceil r/3\rceil}$ maximises the number of $(r,1)$-colourings, completing the proof of Theorem 3.1.3. $\square$

3.4 Concluding remarks

In this chapter we study the Erdős-Rothschild extension for Erdős-Ko-Rado Theorem within various setting in discrete mathematics. We determined the maximum possible number of edge colourings of a hypergraph such that every colour class forms an intersecting hypergraph. Nevertheless, numerous open problems remain.

We could extend Theorem 3.1.1 for $(r,1)$-colourings with $5 \leq r = o\left(\sqrt{\frac{n}{\log n}}\right)$ showing that the extremal families are isomorphic to the family $\bigcup_{i=1}^{\lceil r/3\rceil} \mathcal{T}_{i,i}$, where $\mathcal{T}_{i,i}$ is the coset $\{\sigma \in S_n | \sigma(i) = i\}$. For $r = 4$, our method allows us to show that one of the two families $\mathcal{T}_{1,1} \cup \mathcal{T}_{1,2}$ and $\mathcal{T}_{1,1} \cup \mathcal{T}_{2,2}$ would maximises the number of colourings. The proofs are almost identical to that of Theorem 3.1.3; we leave the details to the readers. The requirement that

$r = o\left(\sqrt{\frac{n}{\log n}}\right)$ seems somewhat artificial; we would expect the same statement to hold for much larger r, say $r = o(n)$, as well.

There is also the question of obtaining sharp dependency between n, k and t in Theorem 3.1.2. For $t = 1$ and $n \geq 3k + 12\log k$, Theorem 3.1.2 shows that k-uniform families on $[n]$ which maximises the number of $(3, 1)$-colourings are stars. New methods will be required to characterise optimal families for the complete range, as the bound on the number of maximal intersecting families is not strong enough to apply Lemma 3.2.1 when $n \leq 3k$. It is worth noting that when $n \geq 2k + 1$, the largest intersecting families are stars. We thus suspect that $n \geq 2k + 1$ may already suffice for the stars to become optimal constructions for the Erdős-Rothschild problem.

We find most exciting the prospect of studying Erdős–Rothschild-type problems in other settings. We hope that further work of this nature will lead to many interesting results and a greater understanding of classical theorems in extremal combinatorics.

Chapter 4

A Density Turán Theorem

4.1 Introduction

Turán-type problems ask for the maximum size of some structures that do not contain a given substructure. In this chapter we consider a multipartite analogue of the problem, suggested by Bollobás (see the discussion after the proof of Theorem VI.2.15 in [12]). Before stating the problem at hand and presenting our contributions, we begin with a brief survey of relevant results.

4.1.1 Background

The fundamental Turán theorem of 1941 [94] completely determined the Turán numbers of a clique: the Turán graph $T_{k-1}(n)$, the complete $(k-1)$-partite graph on n vertices with parts as equal as possible, has the largest number of edges among all K_k-free n-vertex graphs. Thus, we have $\mathrm{ex}(n, K_k) = t_{k-1}(n)$, where $t_{k-1}(n)$ is the number of edges in $T_{k-1}(n)$. This theorem generalises a previous result by Mantel [79] from 1907, which states that $\mathrm{ex}(n, K_3) = \lfloor \frac{n^2}{4} \rfloor$.

A large and important class of graphs for which the Turán numbers are well-understood is formed by *colour-critical graphs*, that is, graphs whose chromatic number can be decreased by removing an edge. Simonovits [92] introduced the stability method to show that $\mathrm{ex}(n, H) = t_{k-1}(n)$ for all $n \geq n_0(H)$ sufficiently large, provided H is a colour-critical graph with $\chi(H) = k$; furthermore, $T_{k-1}(n)$ is the unique extremal graph. As the cliques are colour-critical, Simonovits' theorem implies Turán's theorem for large n.

For general graphs H we still do not know how to compute the Turán numbers $\mathrm{ex}(n, H)$ exactly; but if we are satisfied with an approximate answer the theory becomes quite simple:

it is enough to know the chromatic number of H. The important and deep theorem of Erdős and Stone [40] together with an observation of Erdős and Simonovits [38] shows that $\mathrm{ex}(n,H) = \left(\frac{\chi(H)-2}{\chi(H)-1} + o(1)\right)\frac{n^2}{2}$, where the $o(1)$ term tends to 0 as n tends to infinity. In the literature, this result is usually referred as the Erdős–Stone–Simonovits theorem.

In the years since these seminal theorems appeared, great efforts have been made to extend them, some of which are discussed in Nikiforov's survey [83]. We are particularly interested in the following two extensions.

For every integer $s \geq 2$, let $K_{k-1}(s)$ denote the complete $(k-1)$-partite graph $K_{k-1}(s,\ldots,s)$, and let $K^+_{k-1}(s)$ be the graph obtained from $K_{k-1}(s)$ by adding an edge to the first class. Nikiforov [82] and Erdős [33] (for $k=3$) proved that for all $k \geq 3$ and all sufficiently small $c > 0$, every graph of sufficiently large order n with $t_{k-1}(n)+1$ edges contains not only a K_k but a copy of $K^+_{k-1}\big(\lfloor c\log n\rfloor\big)$. For fixed k, the Erdős-Rényi random graph $G_{n,p}$ shows that the lower bound $c\log n$ on the size of the subgraph in this result is optimal up to a constant factor.

Seeking an extension of Turán's theorem, Erdős [35] asked how many K_k sharing a common edge must exist in a graph on n vertices with $t_{k-1}(n)+1$ edges. Bollobás and Nikiforov [16] sharpened Erdős's result [35] showing that for large enough n, every graph of order n with $t_{k-1}(n)+1$ edges has an edge that is contained in $k^{-k-4}n^{k-2}$ copies of K_k. This result is best possible, up to a poly(k) factor.

In this chapter we shall study analogues of these results for multipartite graphs. For a graph H and an integer $\ell \geq v(H)$, let $d_\ell(H)$ be the minimum real number such that every ℓ-partite graph $G = (V_1 \cup \ldots \cup V_\ell, E)$ with $d(V_i,V_j) := \frac{e(V_i,V_j)}{|V_i||V_j|} > d_\ell(H)$ for all $i \neq j$ contains a copy of H. The problem of determining the exact value of $d_\ell(H)$ was suggested by Bollobás (see the discussion after the proof of Theorem VI.2.15 in [12]). However, it was first studied systematically by Bondy, Shen, Thomassé and Thomassen [17]. Amongst other things Bondy et.al. showed that for every graph H the sequence $d_\ell(H)$ decreases to $\frac{\chi(H)-2}{\chi(H)-1}$ as ℓ tends to infinity. To show the lower bound $d_\ell(H) \geq \frac{\chi(H)-2}{\chi(H)-1}$, they observed that the ℓ-partite graph G obtained from the empty graph on $\{1,\ldots,\ell\}$ by splitting each vertex v of $\{1,\ldots,\ell\}$ into $\chi(H)-1$ vertices $v_1, v_2, \ldots, v_{\chi(H)-1}$, and joining two vertices x_i and y_j if and only if $x \neq y$ and $i \neq j$, has all edge densities equal to $\frac{\chi(H)-2}{\chi(H)-1}$. Since G is $(\chi(H)-1)$-colourable (with vertex classes $V_i = \{v_i : v \in \{1,\ldots,\ell\}\}$ for $1 \leq i \leq \chi(H)-1$), it does not contain a copy of H. For the opposite inequality $\lim\limits_{\ell\to\infty} d_\ell(H) \leq \frac{\chi(H)-2}{\chi(H)-1}$, they used the Erdős–Stone–Simonovits theorem together with an averaging argument.

When $H = K_3$, the aforementioned result of Bondy et. al. [17] implies that $d_\ell(K_3)$ decreases to $\frac{1}{2}$ as ℓ tends to infinity. They also showed that $d_3(K_3) = \frac{-1+\sqrt{5}}{2} \approx 0.61$, $d_4(K_3) >$

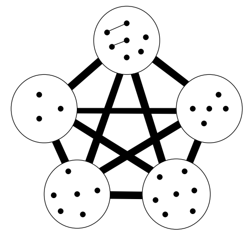

Figure 4.1: An almost colour-critical graph.

0.51, and speculated that $d_\ell(K_3) > \frac{1}{2}$ for all $\ell \geq 3$. Refuting this conjecture, Pfender [85] proved that $d_\ell(K_k) = \frac{k-2}{k-1}$ for large enough ℓ. He also described the family $\mathcal{G}_\ell^k$ of extremal graphs; we shall define this family later in Section 4.2.2.

Theorem 4.1.1 (Pfender [85]). *For every integer $k \geq 3$ there exists a constant $C = C(k)$ such that the following holds for every integer $\ell \geq C$. If $G = (V_1 \cup \ldots \cup V_\ell, E)$ is an ℓ-partite graph with*

$$d(V_i, V_j) \geq \tfrac{k-2}{k-1} \text{ for } i \neq j,$$

then either G contains a K_k or G is isomorphic to a graph in $\mathcal{G}_\ell^k$. In particular, $d_\ell(K_k) = \frac{k-2}{k-1}$ for every $\ell \geq C$.

This theorem can be seen as a multipartite version of the Turán theorem. For an arbitrary graph H, Pfender suggested that $d_\ell(H)$ should be equal to $\frac{\chi(H)-2}{\chi(H)-1}$ for every $\ell \geq \ell_0(H)$ sufficiently large.

4.1.2 Our results

In this chapter we shows that Pfender's suggestion is not quite true. In fact, we characterise those graphs for which the sequence $d_\ell(H)$ is eventually constant, calling them *almost colour-critical.*

Definition 4.1.2. A graph H is called almost colour-critical if there exists a map ϕ from $V(H)$ to $\{1, 2, \ldots, \chi(H) - 1\}$ such that

(i) The induced subgraph of H on $\phi^{-1}(1)$ has maximum degree at most 1,

(ii) For $2 \leq i \leq \chi(H) - 1$, $\phi^{-1}(i)$ is an independent set of H.

In other words, an almost colour-critical graph H has a vertex-colouring with $\chi(H)-1$ colours that is almost proper: all colour classes but one are independent sets, and the exceptional class induces just a matching (see Figure 4.1). For example, cliques, or, more generally colour-critical graphs, are almost colour-critical while the complete k-partite graphs $K_k(s_1, \ldots, s_k)$ are not for every $s_1 \geq 1, s_2 \geq 2, \ldots, s_k \geq 2$.

Our main result shows that almost colour-critical graphs are exactly those for which the sequence $d_\ell(H)$ is eventually constant.

Theorem 4.1.3 ([81])**.** *The following statement holds for every graph H.*

(1) *If H is not almost colour-critical, then $d_\ell(H) \geq \frac{\chi(H)-2}{\chi(H)-1} + \frac{1}{(\chi(H)-1)^2(\ell-1)^2}$ for every $\ell \geq v(H)$.*
(2) *If H is an almost colour-critical graph, then there exists a positive integer $C = C(H)$ so that $d_\ell(H) = \frac{\chi(H)-2}{\chi(H)-1}$ for every $\ell > C$.*

Note that the estimate in the first statement is tight for $H = K_{1,2}$, and the second statement implies Pfender's result since cliques are almost colour-critical. This result can be viewed as a multipartite version of the Simonovits theorem. Since the proof uses the graph removal lemma, the resulting constant $C(H)$ is fairly large.

The rest of the chapter deals with various extensions of Pfender's result. More precisely, we investigate the extensions of Turán's theorem discussed in Section 4.1.1 for balanced multipartite graphs. An ℓ-partite graph G on non-empty independent sets $V_1, \ldots, V_\ell$ is *balanced* if the vertex classes $V_1, \ldots, V_\ell$ are of the same size.

A multipartite version of the extension considered by Nikiforov [82] and Erdős [33] can be stated as follows.

Theorem 4.1.4 ([81])**.** *Let k and ℓ be integers with $k \geq 3$ and $\ell \geq e^{4k^{(k+6)k}}$, and let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph on n vertices such that*

$$d(V_i, V_j) \geq \tfrac{k-2}{k-1} \quad \text{for } i \neq j.$$

Then, either G is isomorphic to a graph in $\mathcal{G}_\ell^k$ or G contains a copy of $K_{k-1}^+\big(\lfloor c \log n \rfloor\big)$, where $c = k^{-(k+6)k}/2$.

For fixed k, the random graph $G_{n,p}$ shows that the lower bound $c \log n$ on the size of the subgraph in this theorem is tight up to a constant factor.

The extension of Turán's theorem studied by Bollobás and Nikiforov [16] has the following multipartite version.

Theorem 4.1.5 ([81])**.** *Let k and ℓ be integers with $k \geq 3$ and $\ell \geq k^{12k}$, and let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph on n vertices such that*

$$d(V_i, V_j) \geq \tfrac{k-2}{k-1} \quad \text{for } i \neq j.$$

Then, G either contains a family of $k^{-2k^2} n^{k-2}$ cliques of order k sharing a common edge or is isomorphic to a graph in $\mathcal{G}_\ell^k$.

With some minor modifications, this result follows from our proof of Theorem 4.1.4. For the sake of clarity we sketch these modifications after detailing the proof of Theorem 4.1.4.

4.1.3 Organisation

The remainder of this chapter is organised as follows. In Section 4.2 we introduce some notation and definitions. In Section 4.3 we extend ideas developed in [85] to prove Theorem 4.1.3. A proof of Theorem 4.1.4 is given in Section 4.4. We sketch how to modify the proof of Theorem 4.1.4 to get Theorem 4.1.5 in Section 4.5, and close with some further remarks and open problems in Section 4.6.

4.2 Preliminaries

4.2.1 Notation

All graphs in this chapter are finite, simple and undirected. Given a graph G, we denote its vertex and edge sets by $V(H)$ and $E(H)$, and the cardinalities of these two sets by $v(H)$ and $e(H)$, respectively. For pairwise disjoint vertex sets $W_1, \ldots, W_r \subseteq V(G)$, we write $G[W_1, \ldots, W_r]$ for the r-colourable graph which can be obtained from $G[W_1 \cup \ldots \cup W_r]$ by deletion of edges in $G[W_i]$ for all $i \leq r$.

Let G be an ℓ-partite graph on non-empty independent sets $V_1, \ldots, V_\ell$. For $X \subseteq V(G)$ and $i \leq \ell$, write $X_i = X \cap V_i$. The edge density between V_i and V_j is $d_{ij} := d(V_i, V_j) := \frac{e(V_i,V_j)}{|V_i||V_j|}$.

For $r \geq 2$ and $t_1 \geq 1, \ldots, t_r \geq 1$, let $K_r(t_1, \ldots, t_r)$ be the complete r-partite graph with classes of sizes $t_1, \ldots, t_r$. If $t_1 = \ldots = t_r = t$, we simply write $K_r(t)$ instead of $K_r(t_1, \ldots, t_r)$. For $r \geq 2$, $s \geq 1$ and $t_1 \geq 2s$, $t_2 \geq 1, \ldots, t_r \geq 1$, we denote by $K_r^{+s}(t_1, \ldots, t_r)$ the graph obtained from $K_r(t_1, \ldots, t_r)$ by adding a matching of size s to the first vertex class. If $s = 1$, we omit the upper index s. In particular, $K_r^{+s}(t)$ is the short form for $K_r^{+s}(t, \ldots, t)$ and $K_r^{+}(t)$ is nothing but $K_r^{+1}(t, \ldots, t)$.

4.2.2 Extremal graphs

In this section we shall recall the definition of the family $\mathcal{G}_\ell^k$ of extremal graphs given by Pfender [85]. For $k \geq 3$ and $\ell \geq (k-1)!$, a graph G is in $\bar{\mathcal{G}}_\ell^k$ if it can be constructed as follows. Let $\{\pi_1, \pi_2, \ldots, \pi_{(k-1)!}\}$ be the set of all permutations of $\{1, \ldots, k-1\}$. For $1 \leq i \leq \ell$ and

$1 \leq s \leq k-1$, pick non-negative integers n_i^s such that

$$n_i^{\pi_i(1)} \geq n_i^{\pi_i(2)} \geq \ldots \geq n_i^{\pi_i(k-1)} \text{ for } 1 \leq i \leq (k-1)!,$$
$$n_i^1 = n_i^2 = \ldots = n_i^{k-1} > 0 \text{ for } (k-1)! < i \leq \ell, \text{ and}$$
$$\sum_s n_i^s > 0 \text{ for } 1 \leq i \leq \ell.$$

Vertex and edge sets of G are defined as (see Figure 4.2)

$$V(G) = \{(i,s,t) : 1 \leq i \leq \ell, 1 \leq s \leq k-1, 1 \leq t \leq n_i^{(s)}\},$$
$$E(G) = \{(i,s,t)(i',s',t') : i \neq i', s \neq s'\}.$$

It is not hard to see that G is an $(k-1)$-colourable ℓ-partite graph with parts $V_i = \{(i,s,t) : 1 \leq s \leq k-1, 1 \leq t \leq n_i^s\}$ for $1 \leq i \leq \ell$, and colour classes $V^{(s)} = \{(i,s,t) : 1 \leq i \leq \ell, 1 \leq t \leq n_i^s\}$ for $1 \leq s \leq k-1$. Moreover, if all n_i^s are equal, we get $d_{ij} = \frac{k-2}{k-1}$ for every $i \neq j$. Note that other weights n_i^s can be used to achieve the inequality $d_{ij} \geq \frac{k-2}{k-1}$ for every $i \neq j$.

Let $\mathcal{G}_\ell^k$ be the family of graphs which can be obtained from graphs in $\bar{\mathcal{G}}_\ell^k$ by removal of some edges in $\{(i,s,t)(i',s',t') : 1 \leq i < i' \leq (k-1)!\}$. The following simple observation by Pfender [85] will be useful for our investigation.

Lemma 4.2.1. *Let $k \geq 3$ and $\ell \geq (k-1)!$ be integers. If $G = (V_1 \cup \ldots \cup V_\ell, E)$ is a $(k-1)$-colourable ℓ-partite graph with $d(V_i, V_j) \geq \frac{k-2}{k-1}$ for $i \neq j$, then it is isomorphic to a graph in $\mathcal{G}_\ell^k$.*

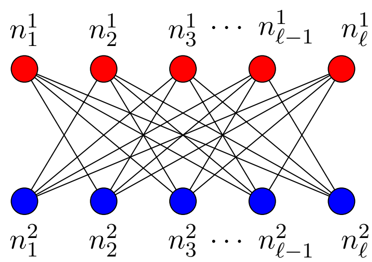

Figure 4.2: A graph in $\bar{\mathcal{G}}_\ell^3$, all edges between different colours in different parts exists.

4.2.3 Infracolourable structures

The following notation will play a key role in our investigation.

Definition 4.2.2. Given a real number $\eta \geq 0$, and integers $k \geq 3$ and $\ell \geq 2$, an (η, k, ℓ)-*infracolourable structure* is an ℓ-partite graph $G = (V_1 \cup \ldots \cup V_\ell, E)$ together with pairs $(D_i^{(s)}, Y_i^{(s)})_{s \leq k-1, i \leq \ell}$ satisfying:

(i) For every $i \leq \ell$, $V_i = \dot{\bigcup}_{s\leq k-1} Y_i^{(s)}$ and $\left|Y_i^{(1)}\right| \geq \left|Y_i^{(2)}\right| \geq \ldots \geq \left|Y_i^{(k-1)}\right|$;

(ii) For every $i \leq \ell$ and every $s \leq k-1$, $D_i^{(s)} \subseteq Y_i^{(s)}$ and $\bigcup_{i\leq \ell} Y_i^{(s)} \setminus D_i^{(s)}$ is an independent set;

(iii) For every $s \leq k-1$, each vertex $v \in \bigcup_{i\leq\ell} D_i^{(s)}$ has at most $\eta \cdot \frac{v(G)}{k-1}$ neighbours in $\bigcup_{i\leq\ell} Y_i^{(s)}$ and at least $3\eta \cdot \frac{v(G)}{k-1}$ non-neighbours in $\bigcup_{i\leq\ell} V_i \setminus Y_i^{(s)}$.

The graph G is called the *base graph* of the infracolourable structure.

Infracolourable structures are useful for us mainly because theirs base graphs break the density conditions in our theorems.

Lemma 4.2.3. *Let η be a positive real number, and let $k \geq 3$ and $\ell \geq 2$ be integers. Suppose that an ℓ-partite graph $G = (V_1 \cup \ldots \cup V_\ell, E)$ together with a system of pairs $(D_i^{(s)}, Y_i^{(s)})_{s\leq k-1, i\leq\ell}$ of vertex sets form an (η, k, ℓ)-infracolourable structure. Then*

$$e(G) \leq \tfrac{k-2}{k-1} \cdot \sum_{i<j} |V_i|\,|V_j|\,.$$

In particular, there exist two different indices i and j such that $d(V_i, V_j) \leq \frac{k-2}{k-1}$. Furthermore, the equality occurs if and only if there exists $i_0 \in \{0, 1, \ldots, \ell\}$ such that $D_i^{(s)} = \emptyset$ for all s and all i, $\left|Y_i^{(s)}\right| = \frac{1}{k-1} \cdot |V_i|$ for all s and all $i \neq i_0$, and $d(Y_i^{(s)}, Y_j^{(t)}) = 1$ for all $s \neq t$ and $i \neq j$.

Proof. It follows from the assumption that

$$\begin{aligned} e(G) &\leq \sum_{\substack{i<j\\ s\neq t}} \left|Y_i^{(s)}\right| \left|Y_j^{(t)}\right| + \left|\bigcup_{i,s} D_i^{(s)}\right| \cdot \left(\eta \cdot \frac{v(G)}{k-1} - \tfrac{1}{2} \cdot 3\eta \cdot \frac{v(G)}{k-1}\right) \\ &\leq \sum_{\substack{i<j\\ s\neq t}} \left|Y_i^{(s)}\right| \left|Y_j^{(t)}\right| = \sum_{i<j} |V_i|\,|V_j| - \sum_{\substack{i<j\\ s\leq k-1}} \left|Y_i^{(s)}\right| \left|Y_j^{(s)}\right| \leq \tfrac{k-2}{k-1} \cdot \sum_{i<j} |V_i|\,|V_j|\,, \end{aligned}$$

where in the last inequality we use Chebyshev's sum inequality. $\square$

To find an infracolourable structure in host graphs we shall need the following technical lemma. It was implicitly stated in [85]. We include a proof here for the sake of completeness.

Lemma 4.2.4. *Let $k \geq 3$ and $\ell \geq 2$ be integers, and let ε be a real number with $0 < \varepsilon < \frac{1}{4}$. Suppose that $G = (V_1 \cup \ldots \cup V_\ell, E)$ is an ℓ-partite graph with $d(V_i, V_j) \geq \frac{k-2}{k-1}$ for all $i \neq j$. Assume that $X_i^{(s)}$ and T_i be subsets of $V(G)$ for $i \leq \ell$ and $s \leq k-1$ with the following three properties:*

(i) *For every $i \leq \ell$, $V_i = X_i^{(1)} \dot{\cup} \ldots \dot{\cup} X_i^{(k-1)} \dot{\cup} T_i$;*

(ii) *For every $i \leq \ell$, $\left|X_i^{(1)}\right| \geq \ldots \geq \left|X_i^{(k-1)}\right|$ and $|T_i| \leq \varepsilon\,|V_i|$;*

(iii) *For every* $s \le k-1$, $\bigcup_{i\le \ell} X_i^{(s)}$ *is an independent set.*

Then there exists a subset $I_0 \in \binom{\mathbb{N}}{k-1}$ *so that* $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm k\sqrt{\varepsilon}\right)|V_i|$ *for* $s \le k-1$ *and* $i \notin I_0$.

Proof. It suffices to show that for each $s \le k-1$ there is at most one index $i \le \ell$ such that $\frac{\left|X_i^{(s)}\right|}{|V_i|} > \frac{1}{k-1} + \sqrt{\varepsilon}$. Assume to the contrary that $\frac{\left|X_i^{(s)}\right|}{|V_i|} \ge \frac{\left|X_j^{(s)}\right|}{|V_j|} > \frac{1}{k-1} + \sqrt{\varepsilon}$ for some s and $i \ne j$. We first prove that $\frac{\left|X_i^{(s)}\right|}{|V_i|} \le 1-\varepsilon$. Otherwise, if $\frac{\left|X_i^{(s)}\right|}{|V_i|} > 1-\varepsilon$, then

$$d(V_i,V_j) \le 1 - \frac{\left|X_i^{(s)}\right|}{|V_i|} \cdot \frac{\left|X_j^{(s)}\right|}{|V_j|} \le 1-(1-\varepsilon)\left(\tfrac{1}{k-1}+\sqrt{\varepsilon}\right) < \tfrac{k-2}{k-1}$$

for $k \ge 3$ and $\varepsilon < \frac{1}{4}$, as $X_i^{(s)} \cup X_j^{(s)}$ is an independent set by (iii). But this contradicts the density condition that $d(V_i,V_j) \ge \frac{k-2}{k-1}$.

We shall get a contradiction by proving that $d(V_i,V_j) < \frac{k-2}{k-1}$. Indeed, we can infer from Chebyschev's sum inequality that

$$\begin{aligned}
d(V_i,V_j) &\overset{(iii)}{\le} 1 - \frac{1}{|V_i|\,|V_j|} \cdot \sum_t \left|X_i^{(t)}\right| \left|X_j^{(t)}\right| \\
&\le 1 - \frac{\left|X_i^{(s)}\right| \left|X_j^{(s)}\right|}{|V_i|\,|V_j|} - \frac{1}{(k-2)\,|V_i|\,|V_j|} \cdot \left(|V_i| - |T_i| - \left|X_i^{(s)}\right|\right)\left(|V_j| - |T_j| - \left|X_j^{(s)}\right|\right) \\
&= 1 - x_i x_j - \tfrac{1}{k-2}\left(1-t_i-x_i\right)\left(1-t_j-x_j\right),
\end{aligned}$$

where $x_i = \frac{\left|X_i^{(s)}\right|}{|V_i|}$, $x_j = \frac{\left|X_j^{(s)}\right|}{|V_j|}$, $t_i = \frac{|T_i|}{|V_i|}$ and $t_j = \frac{|T_j|}{|V_j|}$. Since both x_i and x_j are bounded from below by $\frac{1}{k-1}$, the expression $f(x_i,x_j,t_i,t_j) := 1 - x_i x_j - \frac{1}{k-2}\left(1-t_i-x_i\right)\left(1-t_j-x_j\right)$ is decreasing with respect to both x_i and x_j. Therefore, the density $d(V_i,V_j)$ is bounded from above by

$$f(x_i,x_j,t_i,t_j) \le f\left(\tfrac{1}{k-1}+\sqrt{\varepsilon}, \tfrac{1}{k-1}+\sqrt{\varepsilon}, t_i, t_j\right) \le f\left(\tfrac{1}{k-1}+\sqrt{\varepsilon}, \tfrac{1}{k-1}+\sqrt{\varepsilon}, \varepsilon, \varepsilon\right) < \tfrac{k-2}{k-1},$$

where the second inequality follows from the assumption that $t_i, t_j \in [0,\varepsilon]$. However, this contradicts the assumption that $d(V_i,V_j) \ge \frac{k-2}{k-1}$. □

4.3 Proof of Theorem 4.1.3

In this section we will prove Therem 4.1.3. We begin with a proof of the first assertion.

Proof of Theorem 4.1.3(1). We prove by contradiction. Assume that $d_\ell(H) < \frac{\chi(H)-2}{\chi(H)-1} + \frac{1}{(\chi(H)-1)^2(\ell-1)^2}$. Let $r = \chi(H)-1$, and let $V_1,\ldots,V_\ell$ be ℓ disjoint sets of size $(\ell-1)r$.

For $i \leq \ell$, we partition V_i into r subsets $V_i^{(1)}, \ldots, V_i^{(r)}$ of size $(\ell-1)$ each. We form a complete bipartite graph between $V_i^{(s)}$ and $V_j^{(t)}$ for $i < j$ and $s \neq t$. We then create a perfect matching in $V_1^{(1)} \cup \ldots \cup V_\ell^{(1)}$ such that there is exactly one edge between $V_i^{(1)}$ and $V_j^{(1)}$ for every $i \neq j$. The resulting graph G satisfies

$$d(V_i, V_j) = \frac{\chi(H)-2}{\chi(H)-1} + \frac{1}{(\chi(H)-1)^2(\ell-1)^2} > d_\ell(H) \quad \text{for } i \neq j.$$

Thus, by the definition of $d_\ell(H)$, G must contain a copy of H. From the construction of G, we can see that H is an almost colour-critical graph. This finishes our proof of Theorem 4.1.3(1). $\square$

Remark 4.3.1. The estimate in Theorem 4.1.3(1) is tight for $K_{1,2}$, that is $d_\ell(K_{1,2}) = \frac{1}{(\ell-1)^2}$ for $\ell \geq 3$. Indeed, let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be an ℓ-partite graph with $d(V_i, V_j) > \frac{1}{(\ell-1)^2}$ for every $i \neq j$. We wish to show that G contains a copy of $K_{1,2}$. Suppose to the contrary that G is $K_{1,2}$-free. For $i \neq j$, we write $V_{i,j}$ for the set of vertices in V_i with at least one neighbour in V_j. Since G is $K_{1,2}$-free, we see that

(i) the edges between V_i and V_j form a perfect matching between $V_{i,j}$ and $V_{j,i}$ for every $i \neq j$;

(ii) $V_{i,j}$ and $V_{i,j'}$ are disjoint for all distinct indices i, j and j'.

Notice that $V_{i,j}$ is non-empty for every $i \neq j$ as $d(V_i, V_j) > 0$. Combining this with property (ii), we conclude that

$$|V_i| \geq \sum_{j \in [\ell]\setminus\{i\}} |V_{i,j}| \geq \ell - 1 \text{ for } i \leq \ell. \tag{4.1}$$

Hence

$$\sum_{1 \leq i < j \leq \ell} \left(\frac{|V_{i,j}|}{|V_i|} + \frac{|V_{j,i}|}{|V_j|} \right) = \sum_{1 \leq i \leq \ell} \left(\sum_{j' \neq i} \frac{|V_{i,j'}|}{|V_i|} \right) \leq \ell.$$

Consequently, there exist $1 \leq i < j \leq \ell$ with $\frac{|V_{i,j}|}{|V_i|} + \frac{|V_{j,i}|}{|V_j|} \leq \frac{\ell}{\binom{\ell}{2}} = \frac{2}{\ell-1}$. By appealing to the Cauchy-Schwarz inequality, we thus get $\sqrt{|V_{i,j}|\,|V_{j,i}|} \leq \frac{1}{\ell-1} \cdot \sqrt{|V_i|\,|V_j|}$. This forces

$$d(V_i, V_j) \overset{(i)}{=} \frac{|V_{i,j}|}{|V_i|\,|V_j|} \overset{(i)}{=} \frac{\sqrt{|V_{i,j}|\,|V_{j,i}|}}{|V_i|\,|V_j|} \leq \frac{1}{(\ell-1)\sqrt{|V_i|\,|V_j|}} \overset{(4.1)}{\leq} \frac{1}{(\ell-1)^2},$$

contradicting the assumption that $d(V_i, V_j) > \frac{1}{(\ell-1)^2}$.

To handle the second statement of Theorem 4.1.3, we shall prove a stronger result.

Theorem 4.3.2 ([81])**.** *Let H be an almost colour-critical graph. Then, there exists a constant $C = C(H)$ such that for every integer $\ell > C$, every ℓ-partite graph $G = (V_1 \cup \ldots \cup V_\ell, E)$ with*

$$d(V_i, V_j) > \frac{\chi(H) - 2}{\chi(H) - 1} \quad \textit{for } i \neq j$$

contains a copy of H whose vertices are in different parts of G.

Remark 4.3.3. Suppose that H is almost colour-critical. Let $k = \chi(H)$ and $q = v(H)$. From the definition of almost colour-critical graphs, H is a subgraph $K_{k-1}^{+q}(2q)$. Moreover, it is easy to see that $\chi(K_{k-1}^{+q}(2q)) = k = \chi(H)$ and $K_{k-1}^{+q}(2q)$ is almost colour-critical. Therefore, if Theorem 4.3.2 holds for $K_{k-1}^{+q}(2q)$, it will hold for H as well.

A sketch of the proof. As the proof of Theorem 4.3.2 is quite technique, we begin with a brief outline of the proof. We shall prove the statement by contradiction. Let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be an ℓ-partite graph with $d(V_i, V_j) > \frac{k-2}{k-1}$ for all $i \neq j$, but G does not contain a copy of $H = K_{k-1}^{+q}(2q)$ whose vertices are in different parts of G. We split the proof into two main parts.

- **Step 1 (Nearly spanning induced $(k-1)$-colourable subgraph):** We use the assumption that G does not contain a copy of $K_{k-1}^{+q}(2q)$ whose vertices are in different parts of G, to show that G has few copies of H. Together with a simple consequence of the graph removal lemma and the Erdős-Simonovits stability theorem (see Proposition 4.3.5), this implies that G contains a $(k-1)$-colourable subgraph F' with the partition $V(F') = F_{(1)} \cup \ldots \cup F_{(k-1)}$ such that $\left|F_{(s)}\right| = \left(\frac{1}{k-1} + o(1)\right) v(G)$ and $\deg_{F'}(v, F_{(s)}) = (1 + o(1)) \left|F_{(s)}\right|$ for every index $s \leq k-1$ and vertex $v \in V(F') \setminus F_{(s)}$. Next, we infer from Lemma 4.3.6 that every monochromatic matching in $G[V(F')]$ whose vertices are in different parts of G has size $O_{k,q}(1)$. This would imply that the graph F obtained from $G[V(F')]$ by deleting $O_{k,q}(1)$ parts is $(k-1)$-colourable (see Lemma 4.3.4).

- **Step 2 (Large infracolourable structure):** The induced subgraph F from Step 1 gives rise to a maximum $(k-1)$-cut of G.[1] We then refine the structure of the cut (restrict to a subset of $[\ell]$ if necessary) to obtain a subset $I \subseteq [\ell]$ of size $\Theta(|V(G)|)$ such that $G[\bigcup_{i \in I} V_i]$ is the base graph of an $(\eta, k, |I|)$-infracolourable structure (see Definition 4.2.2). But according to Lemma 4.2.3, this forces $d(V_i, V_j) \leq \frac{k-2}{k-1}$ for some $i, j \in I$, violating the density condition! The first property of an infracolourable structure can be achieved by applying the pigeonhole principle. Lemma 4.3.10 is crucial for getting the second property; its proof relies on Lemma 4.3.6. For the third property, we use Lemma 4.3.9 whose proof follows from Lemmas 4.3.6 and 4.3.8.

[1] A partition of $V(G)$ into $k-1$ disjoint sets so that the number of edges between disjoint parts is maximised.

Our first step in the proof of Theorem 4.3.2 is to show that a counterexample G must contain an induced $(\chi(H)-1)$-colourable subgraph which almost spans $V(G)$. For that we shall need the following stability result.

Lemma 4.3.4. *Given integers $k \geq 3$ and $q \geq 1$ and a real number $0 < \varepsilon < \frac{1}{8k^2q}$, there exists a constant $C = C(k,q,\varepsilon)$ such that the following holds for $\ell \geq C$. Let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph on n vertices with $d(V_i, V_j) \geq \frac{k-2}{k-1}$ for all $i \neq j$. Suppose G contains no copy of $K_{k-1}^{+q}(2q)$ whose vertices lie in different parts of G. Then, G contains an induced $(k-1)$-colourable subgraph F whose vertex classes $X^{(1)}, \ldots, X^{(k-1)}$ satisfy the following properties*

(i) *For $s \leq k-1$, $\left|X^{(s)}\right| = \left(\frac{1}{k-1} \pm \varepsilon\right) n$;*
(ii) *For $s \leq k-1$ and $v \in \bigcup_{t \neq s} X^{(t)}$, $\deg(v, X^{(s)}) \geq \left|X^{(s)}\right| - \varepsilon n$.*

In the proof of Lemma 4.3.4 we shall use the following result whose proof can be found in Section 4.5.

Proposition 4.3.5. *For every graph H and every $\varepsilon > 0$, there exist positive constants $\gamma = \gamma(H,\varepsilon)$ and $C = C(H,\varepsilon)$ such that the following holds for $n \geq C$. Suppose that G is an n-vertex graph with $e(G) \geq \left(\frac{\chi(H)-2}{\chi(H)-1} - \gamma\right)\binom{n}{2}$ containing at most $\gamma n^{v(H)}$ copies of H. Then, G contains a $(\chi(H)-1)$-colourable subgraph of order at least $(1-\varepsilon)n$ and minimum degree at least $\left(\frac{\chi(H)-2}{\chi(H)-1} - \varepsilon\right) n$.*

Another tool that will be used in the proof of Lemma 4.3.4 and Theorem 4.3.2 is an embedding result. Before stating it, we shall introduce the necessary terminology. Let $G[W^{(1)}, \ldots, W^{(r)}]$ be an r-colourable graph such that $W^{(s)} = \dot{\bigcup}_{i \geq 1} W_i^{(s)}$ for every $s \leq r$. We call an embedding $f : K_r(a_1, \ldots, a_r) \to G$ *good* if the sth vertex class of $K_r(a_1, \ldots, a_r)$ is mapped to $W^{(s)}$ for every $s \leq r$, and for each index i there is at most one vertex $v \in K_r(a_1, \ldots, a_r)$ with $f(v) \in \bigcup_{s \leq r} W_i^{(s)}$.

Lemma 4.3.6. *Suppose that $r \geq 2$ and $q \geq 1$ are integers, and let $G[W^{(1)}, \ldots, W^{(r)}]$ be an r-colourable graph which satisfies the following properties*

(i) *For $s \leq r$, $W^{(s)} = \dot{\bigcup}_i W_i^{(s)}$ and $\left|W_i^{(s)}\right| < \frac{1}{2rq} \cdot \left|W^{(s)}\right|$ for all i,*
(ii) *For $s \leq r$ and $v \in \bigcup_{t \neq s} W^{(t)}$, $\deg(v, W^{(s)}) > (1 - \frac{1}{2rq}) \cdot \left|W^{(s)}\right|$.*

Then, for every r-tuple of integers $a_1, \ldots, a_r \in [0,q]$, every good embedding from $K_r(a_1, \ldots, a_r)$ to G can be extended to a good embedding from $K_r(q)$ to G.

Proof. Suppose f is a good embedding from $K_r(a_1, \ldots, a_r)$ to G. To prove the lemma, it suffices to show that f can be extended to a good embedding g from $K_r(a_1, \ldots, a_s + 1, \ldots, a_r)$

to G whenever $a_s \le q-1$. Let v be the vertex of $K_r(a_1,\ldots,a_s+1,\ldots,a_r)$ which is not in $K_r(a_1,\ldots,a_r)$, and let X denote the set of vertices of $K_r(a_1,\ldots,a_r)$ which are not in the sth vertex class. By property (ii), we see that each vertex of X has at most $\frac{1}{2rq}\cdot\left|W^{(s)}\right|$ non-neighbours in $W^{(s)}$, and thus $\left|N(X)\cap W^{(s)}\right| \ge \left|W^{(s)}\right| - |X|\cdot\frac{\left|W^{(s)}\right|}{2rq} \ge \frac{1}{2}\left|W^{(s)}\right|$. Note that, by property (i), each vertex of X can forbid at most $\frac{1}{2rq}\cdot\left|W^{(s)}\right|$ vertices of $W^{(s)}$ from being the image of v. Therefore, the number of possible images of v under g is at least $\left|N(X)\cap W^{(s)}\right| - |X|\cdot\frac{\left|W^{(s)}\right|}{2rq} \ge \frac{1}{2}\left|W^{(s)}\right| - |X|\cdot\frac{\left|W^{(s)}\right|}{2rq} > 0$, where in the last inequality we use the inequality $\left|W^{(s)}\right| > 0$ which is implied by property (i). □

Proof of Lemma 4.3.4. We denote $H = K^{+q}_{k-1}(2q)$, and let

$$\gamma = \gamma_{4.3.5}\left(H, \tfrac{\varepsilon}{2k}\right),\ C = \max\left\{2k^2q^2\gamma^{-1}, 8(k-1)^2q, 4(k-1)q\varepsilon^{-1}, C_{4.3.5}(H, \tfrac{\varepsilon}{2k})\right\}.$$

Because $G = (V_1\cup\ldots\cup V_\ell, E)$ is a balanced ℓ-partite graph on n vertices, we must have

$$|V_1| = |V_2| = \ldots = |V_\ell| = \frac{n}{\ell} := m. \tag{4.2}$$

In the first step, we shall use Proposition 4.3.5 to show that G contains an almost spanning $(k-1)$-colourable subgraph. Indeed, by the choice of C we see that $n \ge \ell \ge C \ge C_{4.3.5}(H, \frac{\varepsilon}{2k})$. Moreover, since G contains no copy of H whose vertices lie in different parts of G, the number of copies of H in G is at most

$$\binom{v(H)}{2}\ell m^2 n^{v(H)-2} < \tfrac{2k^2q^2}{\ell}\cdot(\ell m)^2 n^{v(H)-2} \le \gamma n^{v(H)},$$

since $n = \ell m$ and $\ell \ge C \ge 2k^2q^2\gamma^{-1}$. Also, by the density condition

$$e(G) \ge \binom{\ell}{2}\tfrac{k-2}{k-1}m^2 \overset{(4.2)}{\ge} \left(\tfrac{k-2}{k-1} - \tfrac{1}{\ell}\right)\binom{n}{2} \ge \left(\tfrac{k-2}{k-1} - \gamma\right)\binom{n}{2},$$

assuming $\ell \ge C \ge 2k^2q^2\gamma^{-1}$. Therefore, we can derive from Proposition 4.3.5 that G contains a $(k-1)$-colourable subgraph F' with

$$v(F') \ge (1-\tfrac{\varepsilon}{2k})n \text{ and } \delta(F') \ge \left(\tfrac{k-2}{k-1} - \tfrac{\varepsilon}{2k}\right)n. \tag{4.3}$$

If $W^{(1)},\ldots,W^{(k-1)}$ are vertex classes of F', then (4.3) implies that

$$\left(\tfrac{1}{k-1} - \tfrac{\varepsilon}{2}\right)n \le \left|W^{(s)}\right| \le \left(\tfrac{1}{k-1} + \tfrac{\varepsilon}{2k}\right)n \quad \text{for } s \le k-1. \tag{4.4}$$

In the second step, we shall prove that the induced subgraph $G[V(F')]$ of G does not contain a large monochromatic matching whose vertices are in different parts of G. Indeed, for $s \le k-1$, let $\mathcal{M}_{(s)}$ denote a maximum matching in $G[W^{(s)}]$ whose vertices are in different parts of G, and let K be a subset of $[\ell]$ containing all indices i such that $\bigcup_{s\le k-1}\mathcal{M}_{(s)}$ has a vertex in V_i. The size of K will be bounded from above in terms of k and q.

Claim 4.3.7. $|K| < 2(k-1)q$.

Proof. We prove the claim by contradiction. Suppose that for some $s \le k-1$, $\mathcal{M}_{(s)}$ contains a matching of size q, say $\{x_1x_2, \ldots, x_{2q-1}x_{2q}\}$, . We wish to show that the following two properties holds:

(i) For $t \le k-1$ and $i \le \ell$, $W^{(t)} = W_1^{(t)} \dot{\cup} \ldots \dot{\cup} W_\ell^{(t)}$ and $\left|W_i^{(t)}\right| < \frac{1}{4(k-1)q} \cdot \left|W^{(t)}\right|$;

(ii) For $t \le k-1$ and $v \in V(F') \setminus W^{(t)}$, $\deg_{F'}(v, W^{(t)}) > \left(1 - \frac{1}{4(k-1)q}\right) \cdot \left|W^{(t)}\right|$.

Property (i) follows from the estimate

$$\left|W_i^{(t)}\right| \le |V_i| = \frac{n}{\ell} < \frac{1}{4(k-1)q} \cdot \left(\frac{1}{k-1} - \tfrac{\varepsilon}{2}\right) n \overset{(4.4)}{<} \frac{1}{4(k-1)q} \cdot \left|W^{(t)}\right|$$

for $\ell \ge C \ge 8(k-1)^2 q$ and $\varepsilon < \frac{1}{8k^2 q}$. To prove (ii), assume that $v \in W^{(s)}$ for some $s \ne t$. Because $W^{(s)}$ is an independent set in F', one has $\left|W^{(t)}\right| - d_{F'}(v, W^{(t)}) \le v(F') - \left|W^{(s)}\right| - \deg_{F'}(v)$. Hence by appealing to (4.3) and (4.4), we get

$$\begin{aligned}\left|W^{(t)}\right| - d_{F'}(v, W^{(t)}) &\le n - \left(\frac{1}{k-1} - \tfrac{\varepsilon}{2}\right) n - \left(\frac{k-2}{k-1} - \tfrac{\varepsilon}{2k}\right) n \\ &\le \varepsilon n < \frac{1}{4(k-1)q} \cdot \left(\frac{1}{k-1} - \tfrac{\varepsilon}{2}\right) n \le \frac{1}{4(k-1)q} \cdot \left|W^{(t)}\right|\end{aligned}$$

for $\varepsilon < \frac{1}{8k^2 q}$. This finishes our verification of (i) and (ii).

Finally, properties (i) and (ii) ensure that we can apply Lemma 4.3.6 with $r_{4.3.6} = k-1$ and $q_{4.3.6} = 2q$ to $G[W^{(1)}, \ldots, W^{(r)}]$ to find a copy of $K_{k-1}(2q)$ whose sth vertex class is $\{x_1, \ldots, x_{2q}\}$ and vertices lie in different parts of G. Since $\{x_1, x_2, \ldots, x_{2q-1}x_{2q}\}$ is a matching in G, the graph G contains a desired copy of H, which contradicts our hypothesis. $\square$

To finish the proof, we shall show that G contains an induced subgraph F with the desired properties. For this purpose, we let $X^{(s)} = W^{(s)} \setminus \bigcup_{i \in K} V_i$ for $s \le k-1$. The maximality of $\mathcal{M}_{(s)}$ implies that $X^{(s)}$ is an independent set in G. So the induced subgraph $F = G[X^{(1)} \cup \ldots \cup X^{(k-1)}]$ is $(k-1)$-colourable. What is left is to prove that F has the desired properties. Since $\varepsilon < \frac{1}{8k^2 q}$ and $\ell \ge C \ge 4(k-1)q\varepsilon^{-1}$, we find that

$$v(F) \ge v(F') - \left|\bigcup_{i \in K} V_i\right| \overset{(4.3),\text{Claim 4.3.7}}{\ge} \left(1 - \tfrac{\varepsilon}{2k}\right) n - 2(k-1)q \cdot \frac{n}{\ell} > (1-\varepsilon)n,$$

$$\delta(F) \ge \delta(F') - \left|\bigcup_{i \in K} V_i\right| \overset{(4.3),\text{Claim 4.3.7}}{\ge} \left(\tfrac{k-2}{k-1} - \tfrac{\varepsilon}{2k}\right) n - 2(k-1)q \cdot \frac{n}{\ell} > \left(\tfrac{k-2}{k-1} - \tfrac{\varepsilon}{2}\right) n.$$

Moreover, by (4.4) we see that $\left|X^{(s)}\right| \le \left|W^{(s)}\right| \le \left(\frac{1}{k-1} + \frac{\varepsilon}{2k}\right) n$ for $s \le k-1$, and hence $\left(\frac{1}{k-1} - \frac{\varepsilon}{2}\right) n \le \left|X^{(s)}\right| \le \left(\frac{1}{k-1} + \frac{\varepsilon}{2k}\right) n$ for $s \le k-1$. Therefore, for $s \le k-1$ and $v \in$

$\bigcup_{t\neq s} X^{(t)}$, there are at most $n - \left|X^{(s)}\right| - d_F(v) \leq n - \left(\frac{1}{k-1} - \frac{\varepsilon}{2}\right) n - \left(\frac{k-2}{k-1} - \frac{\varepsilon}{2}\right) n = \varepsilon n$ missing edges in F between v and $X^{(s)}$. This completes our proof of Lemma 4.3.4. $\square$

We also need the following elementary lemma. It is probably well-known, but we could not find a reference. For completeness we include its proof in Section 4.5.

Lemma 4.3.8. *Given integers $r \geq 1$ and $q \geq 2$ and a real number $d \in (0,1)$, there exist an integer $D = D(r,q,d)$ and a positive $\rho = \rho(r,q,d)$ so that the following holds. Suppose that G is an $(r+1)$-colourable graph with vertex classes $U, W_{(1)}, \ldots, W_{(r)}$. If $|U| \geq D$ and $\deg(u, W_{(s)}) \geq d\left|W_{(s)}\right|$ for all $u \in U$ and $s \leq r$, then there is a subset $A \in \binom{U}{q}$ with $\left|N(A) \cap W_{(s)}\right| \geq \rho\left|W_{(s)}\right|$ for $s \leq r$.*

In order to get the third property of an infracolourable structure (see Definition 4.2.2) we shall make use of a consequence of Lemmas 4.3.6 and 4.3.8.

Lemma 4.3.9. *Given integers $k \geq 3$ and $q \geq 1$ and a real number $\eta \in (0,1)$, there exist integers $C = C(k,q,\eta)$ and $D = D(k,q,\eta)$ and a positive $\delta = \delta(k,q,\eta)$ such that the following holds for $\ell \geq C$ and $\varepsilon \in (0,\delta)$. Suppose that $G = (V_1 \cup \ldots \cup V_\ell, E)$ is a balanced ℓ-partite graph containing no copy of $K_k(2q)$ in G whose vertices are in different parts of G. Assume $(X_i^{(s)})_{s\leq k-1, i\leq \ell}$ are vertex sets satisfying:*

(i) *For $i \leq \ell$, $X_i^{(1)}, \ldots, X_i^{(k-1)}$ are disjoint subsets of V_i,*
(ii) *For $i \leq \ell$ and $s \leq k-1$, $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm \varepsilon\right)|V_i|$,*
(iii) *For every $s \leq k-1$ and $v \in \bigcup_{i\leq \ell, t\neq s} X_i^{(t)}$, $\deg(v, \bigcup_{i\leq\ell} X_i^{(s)}) \geq \left|\bigcup_{i\leq\ell} X_i^{(s)}\right| - \varepsilon \cdot v(G)$.*

Let I be the subset of $[\ell]$ consisting of all indices $i \in [\ell]$ such that V_i contains a vertex v with $\deg(v, \bigcup_{j\leq\ell} X_j^{(s)}) \geq \eta \cdot v(G)$ for $s \leq k-1$. Then $|I| \leq D$.

Proof. Let $D = D_{4.3.8}\left(k-1, 2q, \frac{k\eta}{4}\right)$, $C = \max\left\{4kD, 2\eta^{-1}D, \frac{9(k-1)kq}{\rho}\right\}$, $\delta = \min\left\{\frac{1}{4k}, \frac{\rho}{8(k-1)kq}\right\}$, where $\rho = \rho_{4.3.8}\left(k-1, 2q, \frac{k\eta}{4}\right)$. We shall prove the lemma by contradiction. Assume that $|I| \geq D$. Let J be an arbitrary subset of I of size D. By the definition of I, for each index $j \in J$ we can find a vertex $v_j \in V_j$ such that $\deg(v_j, \bigcup_{i\leq\ell} X_i^{(s)}) \geq \eta \cdot v(G)$ for $s \leq k-1$. Let $U = \{v_j : j \in J\}$.

For simplicity of notation, let $X^{(s)} := \bigcup_{i\leq\ell} X_i^{(s)}$ and $W^{(s)} := \bigcup_{i\in[\ell]\setminus J} X_i^{(s)}$ for $s \leq k-1$. Then, property (i) implies that $W^{(1)}, \ldots, W^{(k-1)}$ are disjoint subsets of $V(G)$. By (i) and (ii), we find that

$$\left|W^{(s)}\right| \geq \left(\frac{1}{k-1} - \varepsilon - \frac{D}{\ell}\right) \cdot v(G) \geq \frac{v(G)}{2k} \tag{4.5}$$

for $\varepsilon \le \delta \le \frac{1}{4k}$ and $\ell \ge C \ge 4kD$. Also, (i) and (ii) force $\left|W^{(s)}\right| \le \left(\frac{1}{k-1}+\varepsilon\right) v(G) \le \frac{2v(G)}{k}$, since $\varepsilon \le \delta \le \frac{1}{4k}$. Combining these two inequalities, we conclude that

$$\deg(v, W^{(s)}) \ge \deg(v, X^{(s)}) - \left|\bigcup_{j\in J} V_j\right| \ge \eta \cdot v(G) - D \cdot \frac{v(G)}{\ell} \ge \frac{\eta}{2} \cdot v(G) \ge \frac{k\eta}{4} \cdot \left|W^{(s)}\right|$$

for $v \in U$ and $s \le k-1$, as $\ell \ge 2\eta^{-1}D$. Furthermore, $|U| = D = D_{4.3.8}\left(k-1, 2q, \frac{k\eta}{4}\right)$, by the definition of D. By applying Lemma 4.3.8 to $G[U, W^{(1)}, \ldots, W^{(k-1)}]$ with $r_{4.3.8} = k-1$, $q_{4.3.8} = 2q$ and $d_{4.3.8} = \frac{k\eta}{4}$, we thus obtain a subset $A \in \binom{U}{2q}$ with

$$\left|N(A) \cap W^{(s)}\right| \ge \rho \left|W^{(s)}\right| \quad \text{for } s \le k-1. \tag{4.6}$$

In the rest of the proof we shall use Lemma 4.3.6 to show that $G[N(A)\cap W^{(1)}, \ldots, N(A)\cap W^{(k-1)}]$ contains a copy of $K_{k-1}(2q)$ whose vertices are in different parts of G. Since this copy lies in $N(A)$, together with vertices of A it forms a copy of $K_k(2q)$ whose vertices belong to different parts of G, contradicting the assumption. It remains to verify the assumptions of Lemma 4.3.6. Indeed, for $s \le k-1$, $N(A) \cap W^{(s)}$ does admit the partition

$$N(A) \cap W^{(s)} = \dot{\bigcup}_{j \notin J}\big(N(A) \cap X_j^{(s)}\big). \tag{4.7}$$

Moreover, since $N(A) \cap W^{(s)} \subseteq X^{(s)}$ for $s \le k-1$, we must have, for $s \le k-1$ and $v \in \bigcup_{t\ne s}\big(N(A)\cap W^{(t)}\big)$,

$$\begin{aligned}\left|N(A) \cap W^{(s)}\right| - \deg(v, N(A)\cap W^{(s)}) &\le \left|\bigcup_{i\le \ell} X_i^{(s)}\right| - \deg(v, \bigcup_{i\le\ell} X_i^{(s)}) \overset{(iii)}{\le} \varepsilon \cdot v(G) \\ &\le \tfrac{1}{4(k-1)q} \cdot \rho \cdot \frac{v(G)}{2k} \overset{(4.5),(4.6)}{\le} \tfrac{1}{4(k-1)q} \cdot \left|N(A)\cap W^{(s)}\right|,\end{aligned}$$

assuming $\varepsilon \le \delta \le \frac{\rho}{8(k-1)kq}$. It can be rewritten as

$$\deg(v, N(A)\cap W^{(s)}) \ge \left(1 - \tfrac{1}{4(k-1)q}\right)\left|N(A)\cap W^{(s)}\right| \text{ for } s \le k-1 \text{ and } v \notin \bigcup_{t\ne s}(N(A)\cap W^{(t)}. \tag{4.8}$$

Also, for every $j \notin J$ and $s \le k-1$, we have

$$\left|N(A)\cap X_j^{(s)}\right| \le |V_j| = \tfrac{v(G)}{\ell} < \tfrac{1}{4(k-1)q} \cdot \rho \cdot \frac{v(G)}{2k} \overset{(4.5),(4.6)}{\le} \tfrac{1}{4(k-1)q} \cdot \left|N(A)\cap W^{(s)}\right| \tag{4.9}$$

because $\ell \ge C \ge \frac{9(k-1)kq}{\rho}$. The inequalities (4.7), (4.8) and (4.9) show that we can apply Lemma 4.3.6 to $G[N(A)\cap W^{(1)}, \ldots, N(A)\cap W^{(k-1)}]$ with $r_{4.3.6} = k-1$ and $q_{4.3.6} = 2q$. $\square$

The following consequence of Lemma 4.3.6 will be needed to achieve the second property of an infracolourable structure (see Definition 4.2.2).

Lemma 4.3.10. *Given integers $k \geq 3$ and $q \geq 1$ and a real number $\eta \in \left(\frac{2q-1}{2(k-1)q}, 1\right)$, there exist an integer $C = C(k,q,\eta)$ and a positive $\delta = \delta(k,q,\eta)$ such that the following holds for every integer $\ell \geq C$ and every $\varepsilon \in (0,\delta)$. Let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph containing no copy of $K_{k-1}^{+q}(2q)$ whose vertices are in different parts of G. Assume $(X_i^{(s)}, Y_i^{(s)})_{s \leq k-1, i \leq \ell}$ are pairs of vertex sets satisfying:*

(i) *For $i \leq \ell$ and $s \leq k-1$, $Y_i^{(1)}, \ldots, Y_i^{(k-1)}$ are disjoint subsets of V_i and $X_i^{(s)} \subseteq Y_i^{(s)}$,*
(ii) *For $i \leq \ell$ and $s \leq k-1$, $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm \varepsilon\right)|V_i|$,*
(iii) *For $s \leq k-1$ and $v \in \bigcup_{i \leq \ell, t \neq s} X_i^{(t)}$, $\deg(v, \bigcup_{i \leq \ell} X_i^{(s)}) \geq \left|\bigcup_{i \leq \ell} X_i^{(s)}\right| - \varepsilon \cdot v(G)$.*

For $i \leq \ell$ and $s \leq k-1$, let $B_i^{(s)}$ denote a subset of $Y_i^{(s)}$ consisting of all vertices v with $\deg(v, \bigcup_{j \leq \ell} X_j^{(t)}) < \eta \cdot v(G)$ for some $t \neq s$. For $s \leq k-1$, write $\mathcal{M}_{(s)}$ for a maximal matching in the induced subgraph $G\big[\bigcup_{i \leq \ell} Y_i^{(s)} \setminus B_i^{(s)}\big]$ of G whose vertices are in different parts of G, and set $J = \{j \in [\ell] : V_j$ contains some vertex in $\bigcup_{s \leq k-1} \mathcal{M}_{(s)}\}$. Then, $|J| < 2(k-1)q$.

Proof. Choose
$$C = \frac{4(k-2)}{\eta'} \text{ and } \delta = \min\left\{\frac{q\eta'}{2q-1}, \frac{\eta'}{4(k-2)}\right\}, \text{ where } \eta' = \eta - \frac{2q-1}{2(k-1)q}.$$
Notice that $\eta' > 0$ as $\eta \in \left(\frac{2q-1}{2(k-1)q}, 1\right)$. We prove the statement by contradiction. Suppose that $\mathcal{M}_{(s)}$ contains a matching $\{x_1x_2, \ldots, x_{2q-1}x_{2q}\}$ of size q for some $s \leq k-1$. Let $X^{(t)}$ denote the vertex set $\bigcup_i X_i^{(s)}$ for $s \leq k-1$. For $t \neq s$, define $W_{(t)} = \bigcup_i \big(N(x_1, \ldots, x_{2q}) \cap X_i^{(t)}\big)$. Then property (i) implies that $W^{(1)}, \ldots, W^{(k-1)}$ are disjoint subsets of $V(G)$. We shall apply Lemma 4.3.6 to find a copy of $K_{k-2}(2q)$ in $G\big[W_{(1)}, \ldots, \widehat{W_{(s)}}, \ldots, W_{(k-1)}\big]$ whose vertices are in different parts of G (here $\widehat{W_{(s)}}$ stands for the empty set). Since this copy lies in $N(x_1, \ldots, x_{2q})$ and since $\{x_1x_2, \ldots, x_{2q-1}x_{2q}\}$ is a matching, G contains a copy of $K_{k-1}^{+q}(2q)$ whose vertices belong to different parts of G, which is impossible. The remaining task is thus to verify the assumptions of Lemma 4.3.6. Indeed, from the definition of $W_{(t)}$ we see that, for $t \neq s$,
$$W_{(t)} = \dot{\bigcup_i} \big(N(x_1, \ldots, x_{2q}) \cap X_i^{(t)}\big). \tag{4.10}$$
By the definition of $\mathcal{M}_{(s)}$, we have $\deg(x, X^{(t)}) \geq \eta \cdot v(G)$ for $x \in \{x_1, \ldots, x_{2q}\}$ and $t \neq s$. Hence
$$\begin{aligned}\left|W_{(t)}\right| = \left|N(x_1, \ldots, x_{2q}) \cap X^{(t)}\right| &\geq 2q\eta \cdot v(G) - (2q-1)\left|X^{(t)}\right| \\ &\overset{(ii)}{\geq} 2q\eta \cdot v(G) - (2q-1)\left(\tfrac{1}{k-1} + \varepsilon\right) v(G) \geq q\eta' \cdot v(G) \end{aligned} \tag{4.11}$$
for $\varepsilon \leq \delta \leq \frac{q\eta'}{2q-1}$. Together with the assumption $\ell \geq C = \frac{4(k-2)}{\eta'}$, this inequality implies that, for $i \leq \ell$ and $t \neq s$,
$$\left|N(x_1, \ldots, x_{2q}) \cap X_i^{(t)}\right| \leq |V_i| = \frac{v(G)}{\ell} \leq \tfrac{q\eta'}{4(k-2)q} \cdot v(G) \leq \tfrac{1}{4(k-2)q} \cdot \left|W_{(t)}\right|. \tag{4.12}$$

On the other hand, we can derive from property (iii) that, for $v \in \bigcup_{i\leq \ell, p\notin\{s,t\}} X_i^{(p)}$,

$$\left|W_{(t)}\right| - \deg(v, W_{(t)}) \leq \varepsilon \cdot v(G) \leq \tfrac{q\eta'}{4(k-2)q} \cdot v(G) \overset{(4.11)}{\leq} \tfrac{1}{4(k-2)q} \cdot \left|W_{(t)}\right|, \tag{4.13}$$

assuming $\varepsilon \leq \delta \leq \frac{\eta'}{4(k-2)}$. It follows from (4.10), (4.12) and (4.13) that we can apply Lemma 4.3.6 to $G\big[W_{(1)}, \ldots, \widehat{W_{(s)}}, \ldots, W_{(k-1)}\big]$ with $r_{4.3.6} = k-2$ and $q_{4.3.6} = 2q$. $\square$

We are now ready to prove Theorem 4.3.2.

Proof of Theorem 4.3.2. Let $k = \chi(H)$. If $k = 2$, then H is a matching. The density condition implies that there is at least one edge between any two parts of G. Hence G contains a matching of size $\frac{\ell}{2} \geq e(H)$ whose vertices are in different parts of G. So from now on we can focus on the case when $k \geq 3$. Moreover, as discussed in Remark 4.3.3, we can suppose that $H = K_{k-1}^{+q}(2q)$ for some positive integer q. To prove Theorem 4.3.2, we assume to the contrary that G does not contain a copy of H whose vertices are in different parts of G. Without loss of generality we can suppose that each part of G has exactly m vertices, where m is a sufficiently large integer. Otherwise, multiply each vertex in each part V_i by a factor of $\frac{m}{|V_i|}$, which has no effect on the densities, and creates no copy of H whose vertices lie in different parts of G.

Choose $\ell = \max\{C_{4.3.4}(k, q, \varepsilon), 1/\varepsilon\}$, where $\varepsilon > 0$ is sufficiently small (to be specified later). Let $\ell_1 = \frac{\ell}{2(k-1)!}$, $\ell_2 = \ell_1 - (k-1)$, $\ell_3 = \frac{\ell_2}{(k-1)!}$ and $\ell_4 = \ell_3 - 2(k-1)q - D$, where $D = D_{4.3.9}\left(k, q, \frac{1}{(6q+10)(k-1)(k-1)!}\right)$. Note that the parameters ℓ and ℓ_i both grow as $\Omega(1/\varepsilon)$.

Our goal is to find an infracolourable struture in G. In the first step, we apply Lemma 4.3.4 to G with $k_{4.3.4} = k$, $q_{4.3.4} = q$ and $\varepsilon_{4.3.4} = \varepsilon < \frac{1}{8k^2q}$ to obtain an induced $(k-1)$-colourable subgraph F of G whose vertex classes $X^{(1)}, \ldots, X^{(k-1)}$ satisfy

$$\left|X^{(s)}\right| = \left(\tfrac{1}{k-1} \pm \varepsilon\right) n \text{ for } s \leq k-1, \tag{4.14}$$

$$\deg(v, X^{(s)}) \geq \left|X^{(s)}\right| - \varepsilon n \text{ for } s \leq k-1 \text{ and } v \in \bigcup_{t\neq s} X^{(t)}. \tag{4.15}$$

Let $T = V(G) \setminus V(F)$. The inequality (4.14) implies that $|T| \leq k\varepsilon n$. This forces $|T_i| \leq 2k\varepsilon m$ for at least half of indices $i \leq \ell$. Since $\ell_1 = \frac{\ell}{2(k-1)!}$, by the pigeon hole principle we can relabel the V_i and the $X^{(s)}$ such that $\left|X_i^{(1)}\right| \geq \left|X_i^{(2)}\right| \geq \ldots \geq \left|X_i^{(k-1)}\right|$ and

$$|T_i| \leq 2k\varepsilon m \text{ for } i \leq \ell_1. \tag{4.16}$$

Hence we can apply Lemma 4.2.4 with $\varepsilon_{4.2.4} = 2k\varepsilon < \frac{1}{4}$ to find a subset $I_0 \in \binom{\mathbb{N}}{k-1}$ such that $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm k\sqrt{2k\varepsilon}\right) m$ for $s \leq k-1$ and $i \in [\ell_1] \setminus I_0$. By reordering parts if necessary, we may assume that

$$\left|X_i^{(s)}\right| = \left(\tfrac{1}{k-1} \pm k\sqrt{2k\varepsilon}\right) m \quad \text{for } s \le k-1 \text{ and } i \le \ell_2. \tag{4.17}$$

For $i \le \ell_2$ we shall partition V_i into $k-1$ subsets $Y_i^{(1)}, \dots, Y_i^{(k-1)}$ as follows. A vertex $v \in V_i$ is assigned to $Y_i^{(s)}$ if $\deg\left(v, \bigcup_{j\le \ell_2} X_j^{(s)}\right) = \min_{t\le k-1} \deg\left(v, \bigcup_{j\le\ell_2} X_j^{(t)}\right)$; if there are more than one such index s, arbitrarily choose one of them.

Claim 4.3.11. *$X_i^{(s)} \subseteq Y_i^{(s)} \subseteq X_i^{(s)} \dot\cup T_i$ and $\left|Y_i^{(s)}\right| = \left(\frac{1}{k-1} \pm 2k\sqrt{2k\varepsilon}\right) m$ for $s \le k-1$ and $i \le \ell_2$.*

Proof. Let v be an arbitrary vertex of $X_i^{(s)}$. Since $X^{(s)}$ is an independent set of G, v has no neighbours in $\bigcup_{j\le\ell_2} X_j^{(s)}$. It thus follows from the definition of $Y_i^{(s)}$ that $v \in Y_i^{(s)}$, and so $X_i^{(s)}$ is a subset of $Y_i^{(s)}$. Combining with the fact that $V_i = \left(\dot\bigcup_s X_i^{(s)}\right) \dot\cup T_i = \dot\bigcup_s Y_i^{(s)}$, we conclude that $Y_i^{(s)} \subseteq X_i^{(s)} \dot\cup T_i$ for $i \le \ell_2$ and $s \le k-1$.

As $X_i^{(s)}$ is a subset of $Y_i^{(s)}$, (4.17) tells us that $\left|Y_i^{(s)}\right| \ge \left|X_i^{(s)}\right| \ge \left(\frac{1}{k-1} - k\sqrt{2k\varepsilon}\right) m$ for $i \le \ell_2$ and $s \le k-1$. Using (4.16) and (4.17), we get

$$\left|Y_i^{(s)}\right| \le \left|X_i^{(s)}\right| + |T_i| \le \left(\tfrac{1}{k-1} + k\sqrt{2k\varepsilon} + 2k\varepsilon\right) m \le \left(\tfrac{1}{k-1} + 2k\sqrt{2k\varepsilon}\right) m$$

for $i \le \ell_2$ and $s \le k-1$, where the first inequality holds since $Y_i^{(s)}$ is a subset of $X_i^{(s)} \cup T_i$. $\square$

Let I be the set of all indices $i \in [\ell_2]$ such that there exists a vertex $v_i \in V_i$ with $\deg(v_i, X_1^{(s)} \cup \ldots \cup X_{\ell_2}^{(s)}) \ge \frac{1}{(6q+10)(k-1)!} \cdot \frac{\ell_2 m}{k-1}$ for $s \le k-1$. Below, we show that the size of I is bounded in terms of k and q. We shall see later that this would imply the third property of an infracolourable structure (see Definition 4.2.2).

Claim 4.3.12. *$|I| \le D$.*

Proof. We require ε to be small enough so that $\max\{k\sqrt{2k\varepsilon}, k^k\varepsilon\} < \delta_{4.3.9}(k,q,\eta)$ and $\ell_2 \ge C_{4.3.9}(k,q,\eta)$, where $\eta := \frac{1}{(6q+10)(k-1)(k-1)!}$. By (4.17), $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm k\sqrt{2k\varepsilon}\right) m$ for $s \le k-1$ and $i \le \ell_2$. Moreover, for $s \le k-1$ and $v \in \bigcup_{i\le\ell_2, t\ne s} X_i^{(t)}$, we have

$$\begin{aligned}\deg(v, \bigcup_{i\le\ell_2} X_i^{(s)}) &\ge \left|\bigcup_{i\le\ell_2} X_i^{(s)}\right| + \deg(v, X^{(s)}) - \left|X^{(s)}\right| \\ &\overset{(4.15)}{\ge} \left|\bigcup_{i\le\ell_2} X_i^{(s)}\right| - \varepsilon n \ge \left|\bigcup_{i\le\ell_2} X_i^{(s)}\right| - k^k\varepsilon\ell_2 m.\end{aligned}$$

Therefore, we can apply Lemma 4.3.9 to $G[V_1 \cup \ldots \cup V_{\ell_2}]$ with input $k_{4.3.9} = k$, $q_{4.3.9} = q$ and $\eta_{4.3.9} = \eta$ to conclude that $|I| \le D_{4.3.9}(k,q,\eta) = D$. $\square$

As $\ell_3 = \frac{\ell_2}{(k-1)!}$, by reordering the V_i and $Y^{(s)}$ if necessary we can ensure

$$V_i = \dot{\bigcup_s} Y_i^{(s)} \text{ and } \left|Y_i^{(1)}\right| \geq \left|Y_i^{(2)}\right| \geq \ldots \geq \left|Y_i^{(k-1)}\right| \quad \text{for } i \leq \ell_3. \tag{4.18}$$

For $i \leq \ell_3$ and $s \leq k-1$, let $B_i^{(s)}$ be the set of all vertices $v \in Y_i^{(s)}$ with the property that $\deg(v, X_1^{(t)} \cup \ldots \cup X_{\ell_3}^{(t)}) < \frac{2q}{2q+1} \cdot \frac{\ell_3 m}{k-1}$ for some $t \neq s$. For $s \leq k-1$, let $\mathcal{M}_{(s)}$ denote a maximal matching in $G\big[\bigcup_{i \leq \ell_3} Y_i^{(s)} \setminus B_i^{(s)}\big]$ whose vertices are in different parts of G, and write J for the collection of all indices $j \in [\ell_3]$ so that $\bigcup_{s \leq k-1} \mathcal{M}_{(s)}$ contains some vertex in V_j. We claim that the size of J is bounded from above by a function of k and q. Later, we shall derive from this the second property of an infracolourable structure (see Definition 4.2.2).

Claim 4.3.13. $|J| < 2(k-1)q$.

Proof. We shall apply Lemma 4.3.10 to $G[V_1 \cup \ldots V_{\ell_3}]$ with $k_{4.3.10} = k$, $q_{4.3.10} = q$ and $\eta_{4.3.10} = \frac{2q}{(k-1)(2q+1)}$ to get $|J| < 2(k-1)q$. Note that $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm k\sqrt{2k\varepsilon}\right) m$ for $s \leq k-1$ and $i \leq \ell_3$, by (4.17). Furthermore, for $s \leq k-1$ and $v \in \bigcup_{i \leq \ell_3, t \neq s} X_i^{(t)}$, we have

$$\deg(v, \bigcup_{i \leq \ell_3} X_i^{(s)}) \overset{(4.15)}{\geq} \left|\bigcup_{i \leq \ell_2} X_i^{(s)}\right| - \varepsilon n \geq \left|\bigcup_{i \leq \ell_2} X_i^{(s)}\right| - k^{2k} \varepsilon \ell_3 m.$$

Finally, we can choose ε sufficiently small so that $\max\{k\sqrt{2k\varepsilon}, k^{2k}\varepsilon\} < \delta_{4.3.10}\left(k, q, \frac{2q}{(k-1)(2q+1)}\right)$ and $\ell_3 \geq C_{4.3.10}\left(k, q, \frac{2q}{(k-1)(2q+1)}\right)$. □

From Claims 4.3.12 and 4.3.13 we can assume (relabelling parts once more if necessary) that $\{1, \ldots, \ell_3\} \setminus (I \cup J) = \{1, \ldots, \ell_4\}$. For $i \leq \ell_4$ and $s \leq k-1$, let $D_i^{(s)}$ be the set consisting of all vertices $v \in Y_i^{(s)}$ such that $\deg(v, Y_1^{(t)} \cup \ldots \cup Y_{\ell_4}^{(t)}) < \frac{2q+1}{2q+2} \cdot \frac{\ell_4 m}{k-1}$ for some $t \neq s$.

Claim 4.3.14. *The ℓ_4-partite graph $G[V_1 \cup \ldots \cup V_{\ell_4}]$ together with pairs $(D_i^{(s)}, Y_i^{(s)})_{s \leq k-1, i \leq \ell_4}$ of vertex sets form an $(\frac{1}{6q+9}, k, \ell_4)$-infracolourable structure.*

Proof. We have to verify the following three properties:

(i) For $i \leq \ell_4$, $V_i = \dot{\bigcup}_{s \leq k-1} Y_i^{(s)}$ and $\left|Y_i^{(1)}\right| \geq \left|Y_i^{(2)}\right| \geq \ldots \geq \left|Y_i^{(k-1)}\right|$;

(ii) For $i \leq \ell_4$ and $s \leq k-1$, $D_i^{(s)} \subseteq Y_i^{(s)}$ and $\bigcup_{i \leq \ell_4} Y_i^{(s)} \setminus D_i^{(s)}$ is an independent set;

(iii) For $s \leq k-1$, every vertex $v \in \bigcup_{i \leq \ell_4} D_i^{(s)}$ has at most $\frac{1}{6q+9} \cdot \frac{\ell_4 m}{k-1}$ neighbours in $\bigcup_{i \leq \ell_4} Y_i^{(s)}$ and at least $\frac{1}{2q+3} \cdot \frac{\ell_4 m}{k-1}$ non-neighbours in $\bigcup_{i \leq \ell_4} V_i \setminus Y_i^{(s)}$.

Property (i) follows directly from (4.18). For (ii), we observe that $B_i^{(s)} \subseteq D_i^{(s)}$ for $i \leq \ell_4$ and $s \leq k-1$. We then deduce property (ii) from the maximality of $\mathcal{M}_{(s)}$. For (iii), we consider an arbitrary vertex $v \in \bigcup_{i\le \ell_4} D_i^{(s)}$. Assume to the contrary that $\deg(v, \bigcup_{i\le\ell_4} Y_i^{(s)}) > \frac{1}{6q+9}\cdot\frac{\ell_4 m}{k-1}$. Then, by Claim 4.3.11, we obtain

$$\deg(v, \bigcup_{i\le\ell_4} X_i^{(s)}) \ge \deg(v, \bigcup_{i\le\ell_4} Y_i^{(s)}) - \left|\bigcup_{i\le\ell_4} T_i\right| \overset{(4.16)}{\ge} \tfrac{1}{6q+9}\cdot\tfrac{\ell_4 m}{k-1} - 2k\varepsilon\ell_4 m > \tfrac{1}{(6q+10)(k-1)!}\cdot\tfrac{\ell_2 m}{k-1}$$

for ε sufficiently small. On the other hand, by (ii), we must have $v \in \bigcup_{i\le\ell_4} D_i^{(s)} \subseteq \bigcup_{i\le\ell_4} Y_i^{(s)}$, and so $\deg(v, \bigcup_{i\le\ell_2} X_i^{(t)}) \ge \deg(v, \bigcup_{i\le\ell_2} X_i^{(s)})$ for all $t \le k-1$. Therefore,

$$\deg(v, \bigcup_{i\le\ell_2} X_i^{(t)}) \ge \deg(v, \bigcup_{i\le\ell_2} X_i^{(s)}) \ge \deg(v, \bigcup_{i\le\ell_4} X_i^{(s)}) > \tfrac{1}{(6q+10)(k-1)!}\cdot\tfrac{\ell_2 m}{k-1}$$

for $t \le k-1$, as $v \in \bigcup_{i\le\ell_4} Y_i^{(s)}$. This contradicts the fact that $\{1,\dots,\ell_4\}\cap I = \emptyset$. Finally, by the definition of $\bigcup_{i\le\ell_4} D_i^{(s)}$, there exists $t \ne s$ such that $\deg(v, \bigcup_{i\le\ell_4} Y_i^{(t)}) < \frac{2q+1}{2q+2}\cdot\frac{\ell_4 m}{k-1}$. Consequently, the number of non-neighbours of v in $\bigcup_{i\le\ell_4} Y_i^{(t)}$ is at least

$$\left|\bigcup_{i\le\ell_4} Y_i^{(t)}\right| - \tfrac{2q+1}{2q+2}\cdot\tfrac{\ell_4 m}{k-1} \overset{\text{Claim 4.3.11}}{\ge} \left(\tfrac{1}{k-1} - 2k\sqrt{2k\varepsilon}\right)\ell_4 m - \tfrac{2q+1}{2q+2}\cdot\tfrac{\ell_4 m}{k-1} > \tfrac{1}{2q+3}\cdot\tfrac{\ell_4 m}{k-1},$$

assuming ε is sufficiently small. $\square$

Claim 4.3.14 tells us that $G[V_1\cup\ldots\cup V_{\ell_4}]$ is the base graph of an $(\frac{1}{6q+9}, k, \ell_4)$-infracolourable structure. By appealing to Lemma 4.2.3, we can find two indices $1 \le i < j \le \ell_4$ with $d(V_i, V_j) \le \frac{k-2}{k-1}$, contradicting the assumption that $d(V_i, V_j) > \frac{k-2}{k-1}$. This completes our proof of Theorem 4.3.2. $\square$

4.4 Proof of Theorem 4.1.4

In this section we shall prove a stronger version of Theorem 4.1.4.

Theorem 4.4.1 ([81])**.** *Let k and ℓ be integers with $k \ge 3$ and $\ell \ge e^{2/c}$, where c is a real number with $0 < c \le k^{-(k+6)k}/2$. Suppose that $G = (V_1\cup\ldots\cup V_\ell, E)$ be a balanced ℓ-partite graph on n vertices such that*

$$d(V_i, V_j) \ge \tfrac{k-2}{k-1} \quad \textit{for } i \ne j.$$

Then, G either contains a copy of $K_{k-1}^{+}\big(\lfloor c\log n\rfloor,\dots,\lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$ or is isomorphic to a graph in $\mathcal{G}_\ell^k$.

The idea of the proof is similar to that of Theorem 4.3.2. We assume that G does not contain a copy of $K_{k-1}^{+}\big(\lfloor c\log n\rfloor,\dots,\lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$. We wish to show that G is isomorphic to a graph in the family $\mathcal{G}_\ell^k$. For this purpose, we apply the stability lemma (Lemma 4.4.2) to find an induced $(k-1)$-colourable subgraph of G which almost spans $V(G)$. We then use the embedding lemma (Lemma 4.4.4) showing that G contains a large infracolourable structure. To conclude the proof, we shall use a bootstrapping argument (Lemma 4.4.8) which allows leveraging a weak structure result into a strong structure result.

In the proof of Theorem 4.4.1 we shall need the following stability lemma.

Lemma 4.4.2. *Let k and ℓ be integers with $k \geq 3$ and $\ell \geq e^{2/c}$, where c is a real number with $0 < c \leq k^{-(k+6)k}/2$. Let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph such that $d(V_i, V_j) \geq \frac{k-2}{k-1}$ for $i \neq j$. If G does not contain $K_{k-1}^{+}\big(\lfloor c\log n\rfloor, \ldots, \lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$, then G has an induced $(k-1)$-colourable subgraph F whose vertex classes $X^{(1)}, \ldots, X^{(k-1)}$ satisfy the following properties with $\varepsilon = 4\ell^{-1/2}$*

(i) *For $s \leq k-1$, $\left|X^{(s)}\right| = \left(\frac{1}{k-1} \pm k\varepsilon\right) n$;*
(ii) *For $s \leq k-1$ and $v \in \bigcup_{t\neq s} X^{(t)}$, $\deg(v, X^{(s)}) \geq \left|X^{(s)}\right| - k\varepsilon n$.*

To prove the above statement we need a stability lemma of Nikiforov [82, Theorem 3].

Lemma 4.4.3. *Let $k \geq 3$ be an integer, and let c and δ be positive real numbers with $c < k^{-(k+6)k}/2$ and $\delta < \frac{1}{8k^8}$. Suppose that G is a graph of order $n \geq e^{2/c}$ with $e(G) \geq \left(\frac{k-2}{k-1} - \delta\right)\binom{n}{2}$. If G has no copy of $K_{k-1}^{+}\big(\lfloor c\log n\rfloor, \ldots, \lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$, then G contains an induced $(k-1)$-colourable subgraph F of order $v(F) \geq (1-2\sqrt{\delta})n$ and minimum degree $\delta(F) \geq \left(\frac{k-2}{k-1} - 4\sqrt{\delta}\right) n$.*

Proof of Lemma 4.4.2. By the assumption, $|V_1| = \ldots = |V_\ell| = \frac{n}{\ell} := m$. Together with the density condition, we conclude that $e(G) \geq \binom{\ell}{2}\frac{k-2}{k-1}m^2 \geq \left(\frac{k-2}{k-1} - \frac{1}{\ell}\right)\frac{(\ell m)^2}{2} = \left(\frac{k-2}{k-1} - \frac{1}{\ell}\right)\frac{n^2}{2}$. Notice that $c \leq k^{-(k+6)k}$, $\frac{1}{\ell} < \frac{1}{8k^8}$ and $n \geq e^{2/c}$. Thus, by applying Lemma 4.4.2 to G with $\delta_{4.4.2} = \frac{1}{\ell}$ we obtain an $(k-1)$-colourable induced subgraph $F = G[X^{(1)} \cup \ldots \cup X^{(k-1)}]$ of G with $v(F) > (1-\varepsilon)n$ and $\delta(F) \geq \left(\frac{k-2}{k-1} - \varepsilon\right)n$. Since $\delta(F) \geq \left(\frac{k-2}{k-1} - \varepsilon\right)n$ and since $X^{(s)}$ is an independent set, we must have

$$\left|X^{(s)}\right| \leq n - \delta(F) \leq \left(\frac{1}{k-1} + \varepsilon\right) n$$

for $s \leq k-1$. This implies that

$$\left|X^{(s)}\right| \geq v(F) - (k-2)\left(\frac{1}{k-1} + \varepsilon\right) n \geq (1-\varepsilon)n - (k-2)\left(\frac{1}{k-1} + \varepsilon\right) n = \left(\frac{1}{k-1} - (k-1)\varepsilon\right) n$$

for $s \leq k-1$. Therefore, for $s \leq k-1$ and $v \in \bigcup_{t\neq s} X^{(t)}$, the number of non-neighbours of v in $X^{(s)}$ is at most

$$n - \left|X^{(s)}\right| - d_F(v) \leq n - \left(\frac{1}{k-1} - (k-1)\varepsilon\right) n - \left(\frac{k-2}{k-1} - \varepsilon\right) n = k\varepsilon n,$$

as desired. $\square$

The next ingredient we need is an embedding result.

Lemma 4.4.4. *Let $r \geq 2$ be an integer, and let G be an r-colourable graph with vertex classes $W_{(1)}, \ldots, W_{(r)}$ of the same size h. Suppose that $\deg(w, W_{(s)}) \geq \left(1 - \frac{1}{r^2}\right)h$ for $s \leq r$ and $w \in \bigcup_{t \neq s} W_{(t)}$. Then*

(1) *G contains at least $\frac{1}{2}h^r$ copies of K_r,*
(2) *For $\alpha \in (0, \frac{1}{4})$ and $s \leq r$, G contains a copy of $K_r(\lfloor \alpha^r \log h \rfloor, \ldots, \lfloor \alpha^r \log h \rfloor, \lfloor h^{1-\alpha^{r-1}} \rfloor)$ whose sth vertex class is a subset of $W_{(s)}$.*

The proof of the above lemma requires a simple result of Nikiforov [82, Lemma 5].

Lemma 4.4.5. *Let $r \geq 2$ be an integer, and let α be a real number in $(0, \frac{1}{4})$. Suppose that $B[U, W]$ is a bipartite graph with $|U| = p$ and $|W| = q$. If $p \geq 4\lfloor \alpha^r \log q \rfloor$ and $e(B[U, W]) \geq \frac{1}{2}pq$, then $B[U, W]$ contains the complete bipartite graph $K(a, b)$ with $a = \lfloor \alpha^r \log q \rfloor$ and $b = \lfloor q^{1-\alpha^{r-1}} \rfloor$.*

Proof of Lemma 4.4.4. (1) Let $w_s \in W_{(s)}$ for $s = 1, \ldots, r$. Observe that $\{w_1, \ldots, w_r\}$ forms a clique of G if and only if $w_s \in N(w_1, \ldots, w_{s-1}) \cap W_{(s)}$ for $s = 2, \ldots, r$. In addition, $\left|N(w_1, \ldots, w_{s-1}) \cap W_{(s)}\right| \geq h - (s-1) \cdot \frac{h}{r^2}$. Thus, we can bound the number of copies of K_r in G from below by

$$h^r \cdot \prod_{s=1}^{r} \left(1 - \tfrac{s-1}{r^2}\right) \geq h^r \cdot \left(1 - \sum_{s=1}^{r} \tfrac{s-1}{r^2}\right) = \tfrac{r+1}{2r} \cdot h^r > \tfrac{1}{2}h^r.$$

(2) We proceed by induction on r. The base case $r = 2$ follows from the first assertion and Lemma 4.4.5. For the induction step, assume that $r > 2$. The induction hypothesis implies that $G[W_{(1)} \cup \ldots \cup W_{(r-1)}]$ contains a copy of $K_{r-1}(m)$ with $m = \lfloor \alpha^{r-1} \log h \rfloor$. Let U denote a set of m disjoint copies of K_{r-1} in $K_{r-1}(m)$. Define a bipartite graph $B[U, W_{(r)}]$ with vertex classes U and $W_{(r)}$, joining $R \in U$ to $w \in W_{(r)}$ if $R \cup \{w\}$ is a clique. We see that $|U| = m$ and $\left|W_{(r)}\right| = h$. Since $0 < \alpha < 1/4$, we have $m = \lfloor \alpha^{r-1} \log h \rfloor \geq \lfloor 4\alpha^r \log h \rfloor \geq 4\lfloor \alpha^r \log h \rfloor$. Furthermore, every vertex of U has at least $h - r \cdot \frac{h}{r^2} \geq h/2$ neighbours in $W_{(r)}$. Hence $e(B[U, W_{(r)}]) \geq mh/2$. The assertion then follows from the base case $r = 2$. □

In order to find a large infracolourable structure in G we shall use the following consequence of Lemma 4.4.4.

Lemma 4.4.6. *Let $k \geq 3$ and $\ell \geq 2$ be integers, and let ε and α be positive real numbers with $\varepsilon < 10^{-2}k^{-k}$ and $\alpha < \frac{1}{4}$. Suppose that $G = (V_1 \cup \ldots \cup V_\ell, E)$ is a balanced ℓ-partite graph containing no copy of $K_{k-1}^{+}\left(\lfloor \alpha^{k-1} \log(p) \rfloor, \ldots, \lfloor \alpha^{k-1} \log(p) \rfloor, \lfloor p^{1-\alpha^{k-2}} \rfloor\right)$, where $p = \frac{1}{16(k-1)(k-1)!} \cdot v(G)$. Assume that $(X_i^{(s)})_{s \leq k-1, i \leq \ell}$ are vertex sets so that*

(i) *For $i \le \ell$, $X_i^{(1)}, \ldots, X_i^{(k-1)}$ are disjoint subsets of V_i;*

(ii) *For $s \le k-1$ and $v \in \bigcup_{i \le \ell, t \ne s} X_i^{(t)}$, $\deg(v, \bigcup_{i \le \ell} X_i^{(s)}) \ge \left|\bigcup_{i \le \ell} X_i^{(s)}\right| - \varepsilon \cdot v(G)$.*

Then, there are no vertices $v \in V(G)$ such that $\deg(v, \bigcup_{i \le \ell} X_i^{(s)}) \ge p$ for all $s \le k-1$.

Proof. Suppose for the contradiction that there is $v \in V(G)$ with $\deg(v, \bigcup_i X_i^{(s)}) \ge p$ for all $s \le k-1$. Then, for $s \le k-1$ there exists a subset

$$W_{(s)} \subseteq N(v) \cap \left(\bigcup_i X_i^{(s)}\right) \text{ with } \left|W_{(s)}\right| = p. \tag{4.19}$$

By property (i), $W_{(1)}, \ldots, W_{(k-1)}$ are disjoint subsets of $V(G)$. On the other hand, property (ii) shows that for all $s \le k-1$ and $v \in \bigcup_{t \ne s} W_{(t)}$ one has

$$\begin{aligned} \deg(v, W_{(s)}) &\ge \left|W_{(s)}\right| - \varepsilon \cdot v(G) \ge \left|W_{(s)}\right| - \frac{1}{(k-1)^2} \cdot \frac{1}{16(k-1)(k-1)!} \cdot v(G) \\ &= \left(1 - \frac{1}{(k-1)^2}\right) \left|W_{(s)}\right|, \end{aligned} \tag{4.20}$$

as $\varepsilon < 10^{-2} k^{-k}$. Finally, it follows from (4.19) and (4.20) that we can apply Lemma 4.4.4(2) to the graph $G[W_{(1)}, \ldots, W_{(k-1)}]$ with $r_{4.4.4} = k-1$, $h_{4.4.4} = p$ and $\alpha_{4.4.4} = \alpha$ to find a copy of $K_{k-1}\big(\lfloor \alpha^{k-1} \log(p) \rfloor, \ldots, \lfloor \alpha^{k-1} \log(p) \rfloor, \lfloor p^{1-\alpha^{k-2}} \rfloor\big)$. Since $W_{(1)} \cup \ldots \cup W_{(k-1)}$ lies in the neighbour of v, G contains a copy of $K_{k-1}^{+}\big(\lfloor \alpha^{k-1} \log(p) \rfloor, \ldots, \lfloor \alpha^{k-1} \log(p) \rfloor, \lfloor p^{1-\alpha^{k-2}} \rfloor\big)$, which contradicts our assumption. □

To find a large infracolourable structure in G we also require the following consequence of Lemma 4.4.4.

Lemma 4.4.7. *Let $k \ge 3$ and $\ell \ge 2$ be integers, and let ε and α be positive real numbers with $\varepsilon < \frac{1}{12k^3}$ and $\alpha < \frac{1}{4}$. Let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph containing no copy of $K_{k-1}^{+}\big(\lfloor \alpha^{k-1} \log(p) \rfloor, \ldots, \lfloor \alpha^{k-1} \log(p) \rfloor, \lfloor p^{1-\alpha^{k-2}} \rfloor\big)$, where $p = \frac{1}{4(k-1)} \cdot v(G)$. Suppose $(X_i^{(s)}, Y_i^{(s)})_{s \le k-1, i \le \ell}$ are pairs of vertex sets which satisfy*

(i) *For every $i \le \ell$ and $s \le k-1$, $Y_i^{(1)}, \ldots, Y_i^{(k-1)}$ are disjoint subsets of V_i and $X_i^{(s)} \subseteq Y_i^{(s)}$;*

(ii) *For $i \le \ell$ and $s \le k-1$, $\left|X_i^{(s)}\right| = \left(\frac{1}{k-1} \pm \varepsilon\right) |V_i|$;*

(iii) *For $s \le k-1$ and $v \in \bigcup_{i \le \ell, t \ne s} X_i^{(t)}$, $\deg(v, \bigcup_{i \le \ell} X_i^{(s)}) \ge \left|\bigcup_{i \le \ell} X_i^{(s)}\right| - \varepsilon \cdot v(G)$.*

For $i \le \ell$ and $s \le k-1$, let $B_i^{(s)}$ stands for a subset of $Y_i^{(s)}$ consisting of all vertices v with $\deg(v, \bigcup_{j \le \ell} X_j^{(t)}) < \frac{2}{3(k-1)} \cdot v(G)$ for some $t \ne s$. Then, for $s \le k-1$, $\bigcup_{i \le \ell} Y_i^{(s)} \setminus B_i^{(s)}$ is an independent set of G.

Proof. We prove by contradiction. Suppose that there exists an edge $\{x, y\} \in E(G)$ with $x, y \in \bigcup_i Y_i^{(s)} \setminus B_i^{(s)}$. Let $t \ne s$. By the definition of $\bigcup_i B_i^{(s)}$, both $\deg(x, \bigcup_i X_i^{(t)})$ and

$\deg(y, \bigcup_i X_i^{(t)})$ are at least $\frac{2}{3(k-1)} \cdot v(G)$. Hence

$$\left|N(x,y) \cap \bigcup_i X_i^{(t)}\right| \geq \deg(x, \bigcup_i X_i^{(t)}) + \deg(y, \bigcup_i X_i^{(t)}) - \left|\bigcup_i X_i^{(t)}\right|$$
$$\overset{(ii)}{\geq} \frac{4}{3(k-1)} \cdot v(G) - \left(\frac{1}{k-1} + \varepsilon\right) \cdot v(G) \geq \frac{1}{4(k-1)} \cdot v(G),$$

as $\varepsilon < \frac{1}{12k^3}$. It means that there is a subset

$$W_{(t)} \subseteq N(x,y) \cap \bigcup_{i \leq \ell_3} X_i^{(t)} \text{ with } \left|W_{(t)}\right| = \tfrac{1}{4(k-1)} \cdot v(G).$$

On the other hand, it follows from property (ii) that $\left|\bigcup_i X_i^{(s)}\right| \geq \left(\frac{1}{k-1} - \varepsilon\right) v(G) > \frac{1}{4(k-1)} \cdot v(G)$ for $0 < \varepsilon < \frac{1}{12k^3}$, and so there exists a subset

$$W_{(s)} \subseteq \bigcup_i X_i^{(s)} \text{ with } \left|W_{(s)}\right| = \tfrac{1}{4(k-1)} \cdot v(G).$$

Analysis similar to that in the proof of Lemma 4.4.6 shows that $G[W_{(1)}, \ldots, W_{(k-1)}]$ must contain a copy of $K_{k-1}\big(\lfloor \alpha^{k-1}\log(p)\rfloor, \ldots, \lfloor \alpha^{k-1}\log(p)\rfloor, \lfloor p^{1-\alpha^{k-2}}\rfloor\big)$ whose sth vertex class is of size $\lfloor \alpha^{k-1}\log(p)\rfloor$. Adding back vertices x and y to this class one gets a supgraph of the graph $K^+_{k-1}\big(\lfloor \alpha^{k-1}\log(p)\rfloor, \ldots, \lfloor \alpha^{k-1}\log(p)\rfloor, \lfloor p^{1-\alpha^{k-2}}\rfloor\big)$, contradicting the hypothesis. □

The last component of the proof is a bootstrapping argument which allows us to leverage a weak structure result into a strong structure result. Roughly speaking, it says that if G contains an $\tilde{\ell}$-partite subgraph which is in $\mathcal{G}^k_{\tilde{\ell}}$, then G must belong to $\mathcal{G}^k_\ell$.

Lemma 4.4.8. *Let $k \geq 3$ be an integer, and let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be an ℓ-partite graph with $|V_1| = \ldots = |V_\ell| = m$ and $d(V_i, V_j) \geq \frac{k-2}{k-1}$ for all $i \neq j$. Suppose that there exist an integer $\tilde{\ell}$ and disjoint subsets $Y_i^{(1)}, \ldots, Y_i^{(k-1)}$ of V_i for $1 \leq i \leq \tilde{\ell}$ so that $\left|Y_i^{(s)}\right| = \frac{m}{k-1}$ and $d(Y_i^{(s)}, Y_j^{(t)}) = 1$ for all $i \neq j$ and $s \neq t$. If G does not contain a copy of $K^+_{k-1}\big(\frac{\tilde{\ell}m}{32k^2}\big)$, then G is isomorphic to a graph in the family $\mathcal{G}^k_\ell$.*

Proof. We wish to show that G is isomorphic to a graph in $\mathcal{G}^k_\ell$. According to Lemma 4.2.1, it suffices to prove G is $(k-1)$-colourable. By the assumption, we have

$$\left|Y_i^{(s)}\right| = \tfrac{m}{k-1},\ d(Y_i^{(s)}, Y_j^{(t)}) = 1 \text{ for } s \neq t \text{ and } 1 \leq i < j \leq \tilde{\ell}. \tag{4.21}$$

We shall show that for $v \in V(G) \setminus \big(V_1 \cup \ldots \cup V_{\tilde{\ell}}\big)$ there does not exist $s \leq k-1$ with

$$\deg\big(v, Y_1^{(s)} \cup \ldots \cup Y_{\tilde{\ell}}^{(s)}\big) \geq 1,\ \deg\big(v, Y_1^{(t)} \cup \ldots \cup Y_{\tilde{\ell}}^{(t)}\big) \geq \frac{\tilde{\ell}m}{2k} \text{ for all } t \neq s. \tag{4.22}$$

We prove by contradiction. Suppose that (4.22) holds. We can pick an index $i_0 \in \{1, 2, \ldots, \tilde{\ell}\}$ with $N(v) \cap Y_{i_0}^{(s)} \neq \emptyset$ whose existence is guaranteed by (4.22). We then arbitrarily add other indices to get a subset $I_{(s)} \subset \{1, \ldots, \tilde{\ell}\}$ of size $\frac{\tilde{\ell}}{8k}$. It follows from (4.21) and (4.22) that for each $t \neq s$, there are at least $\frac{\tilde{\ell}}{4}$ indices $i \leq \tilde{\ell}$ with $\deg(v, Y_i^{(t)}) \geq \frac{m}{4k}$. Hence we can find $k-1$ disjoint subsets $I_{(1)}, \ldots, I_{(k-1)}$ of size $\frac{\tilde{\ell}}{8k}$ of $\{1, \ldots, \tilde{\ell}\}$ with the property that $\deg(v, Y_i^{(t)}) \geq \frac{m}{4k}$ for all $t \neq s$ and $i \in I_{(t)}$. By (4.21), $G\big[\bigcup_{i \in I_{(1)}} Y_i^{(1)}, \ldots, \bigcup_{i \in I_{(k-1)}} Y_i^{(k-1)}\big]$ is a complete $(k-1)$-partite graph. In addition, we have $\Big|N(v) \cap \bigcup_{i \in I_{(s)}} Y_i^{(s)}\Big| \geq \Big|N(v) \cap Y_{i_0}^{(s)}\Big| > 0$ and

$$\left|N(v) \cap \bigcup_{i \in I_{(t)}} Y_i^{(t)}\right| = \sum_{i \in I_{(t)}} \deg(v, Y_i^{(t)}) \geq \left|I_{(t)}\right| \cdot \frac{m}{4k} = \frac{\tilde{\ell} m}{32k^2} \quad \text{for } t \neq s.$$

Therefore, by adding v to the sth part of $G\big[\bigcup_{i \in I_{(1)}} Y_i^{(1)}, \ldots, \bigcup_{i \in I_{(k-1)}} Y_i^{(k-1)}\big]$ we get a supergraph of $K_{k-1}^{+}\big(\frac{\tilde{\ell} m}{32k^2}\big)$ in G, contradicting our assumption.

We can infer from (4.22) that $\deg(v, V_1 \cup \ldots \cup V_{\tilde{\ell}}) \leq \frac{k-2}{k-1} \cdot \tilde{\ell} m$ for all $v \in V(G) \setminus \big(V_1 \cup \ldots \cup V_{\tilde{\ell}}\big)$. By the density condition, equality must hold. Again (4.22) shows that for each $v \in V(G) \setminus \big(V_1 \cup \ldots \cup V_{\tilde{\ell}}\big)$,

$$N(v) \cap \big(V_1 \cup \ldots \cup V_{\tilde{\ell}}\big) = \bigcup_{i \leq \tilde{\ell}} V_i \setminus Y_i^{(s)} \quad \text{for some } s \leq k-1. \tag{4.23}$$

If $v \in V_i$ for some $i > \tilde{\ell}$, then we assign v to $Z_i^{(s)}$. For $i \leq \tilde{\ell}$ we let $Z_i^{(s)} = Y_i^{(s)}$ for $s \leq k-1$. If we denote $Z^{(s)} = \dot{\bigcup}_i Z_i^{(s)}$ for $s \leq k-1$, then $V = \dot{\bigcup}_s Z^{(s)}$. To prove G is $(k-1)$-colourable, it is enough to show that $Z^{(1)}, \ldots, Z^{(k-1)}$ are independent sets. Suppose to the contrary that for some $s \leq k-1$, $Z^{(s)}$ contains an edge $\{u, v\}$ with $u \in Z_{i_1}^{(s)}$ and $v \in Z_{i_2}^{(s)}$. We can easily find $k-1$ disjoint subsets $J_{(1)}, \ldots, J_{(k-1)}$ of size $\frac{\tilde{\ell}}{2(k-1)}$ of $[\tilde{\ell}] \setminus \{i_1, i_2\}$. Let $W^{(s)} = \{u, v\} \cup \Big(\bigcup_{i \in J_{(s)}} Y_i^{(s)}\Big)$ and $W^{(t)} = \bigcup_{i \in J_{(t)}} Y_i^{(t)}$ for $t \neq s$. It follows from (4.21) and (4.23) that $G[W^{(1)}, \ldots, W^{(k-1)}]$ is a complete $(k-1)$-colourable graph with $\left|W^{(t)}\right| \geq \frac{\tilde{\ell}}{2(k-1)} \cdot \frac{m}{k-1} > \frac{\tilde{\ell} m}{32k^2}$ for $t \leq k-1$. Combining this with the assumption that $\{u, v\} \in E(G)$, we conclude that G contains a copy of $K_{k-1}^{+}\big(\frac{\tilde{\ell} m}{32k^2}\big)$, a contradiction. □

We now have all the necessary tools to prove Theorem 4.4.1.

Proof of Theorem 4.4.1. For convenience, $H = K_{k-1}^{+}\big(\lfloor c \log n \rfloor, \ldots, \lfloor c \log n \rfloor, \lfloor n^{1-2\sqrt{c}} \rfloor\big)$ and $H^{-} = K_{k-1}\big(\lfloor c \log n \rfloor, \ldots, \lfloor c \log n \rfloor, \lfloor n^{1-2\sqrt{c}} \rfloor\big)$. Suppose G has no copy of H. We wish to show that G is isomorphic to a graph in $\mathcal{G}_{\ell}^{k}$. Since G is a balanced ℓ-partite graph on n vertices, each partition set of G has size $n/\ell := m$. Let $\varepsilon = 4\ell^{-1/2}$, $\ell_1 = \frac{\ell}{2(k-1)!} - (k-1)$, $\ell_2 = \frac{\ell_2}{(k-1)!}$ and $\ell_3 = \ell_2 - 1$.

By Lemma 4.4.2, G must contain an induced $(k-1)$-colourable subgraph F whose vertex classes $X^{(1)}, \dots, X^{(k-1)}$ satisfy

$$\left|X^{(s)}\right| = \left(\tfrac{1}{k-1} \pm k\varepsilon\right) n \text{ for } s \leq k-1, \tag{4.24}$$

$$\deg(v, X^{(s)}) \geq \left|X^{(s)}\right| - k\varepsilon n \text{ for } s \leq k-1 \text{ and } v \in \bigcup_{t \neq s} X^{(t)}. \tag{4.25}$$

Let $T = V(G) \setminus V(F)$. As in the proof of Theorem 4.3.2, by relabelling parts we can assume that

$$|T_i| \leq 2k^2 \varepsilon m, \text{ and } \left|X_i^{(s)}\right| = \left(\tfrac{1}{k-1} \pm 2k^2\sqrt{\varepsilon}\right) m \quad \text{for } i \leq \ell_1 \text{ and } s \leq k-1. \tag{4.26}$$

For $i \leq \ell_1$ we shall partition V_i into $k-1$ subsets as follows. A vertex $v \in V_i$ is assigned to $Y_i^{(s)}$ if $\deg\left(v, \bigcup_{j \leq \ell_1} X_j^{(s)}\right) = \min_{t \leq k-1} \deg\left(v, \bigcup_{j \leq \ell_1} X_j^{(t)}\right)$; if there are more than one such index s, arbitrarily pick one of them.

Claim 4.4.9. *$X_i^{(s)} \subseteq Y_i^{(s)} \subseteq X_i^{(s)} \dot{\cup} T_i$ for $i \leq \ell_1$ and $s \leq k-1$.*

Proof. Because $X^{(s)}$ is an independent set in G, every vertex in $X_i^{(s)}$ has no neighbours in $\bigcup_{j \leq \ell_1} X_j^{(s)}$, and so $X_i^{(s)}$ is a subset of $Y_i^{(s)}$. Since $V_i = \left(\dot{\bigcup}_s X_i^{(s)}\right) \dot{\cup} T_i = \dot{\bigcup}_s Y_i^{(s)}$ and $X_i^{(s)} \subseteq Y_i^{(s)}$ for $i \leq \ell_1$ and $s \leq k-1$, the inclusion relation $Y_i^{(s)} \subseteq X_i^{(s)} \dot{\cup} T_i$ holds for $i \leq \ell_1$ and $s \leq k-1$. □

We proceed by showing that $\bigcup_{i \leq \ell_1} V_i$ does not contain a vertex which has relatively large degree to $\bigcup_{i \leq \ell_1} Y_i^{(s)}$ for all $s \leq k-1$.

Claim 4.4.10. *There are no vertices $v \in \bigcup_{i \leq \ell_1} V_i$ with $\deg(v, \bigcup_{i \leq \ell_1} Y_i^{(s)}) \geq \frac{1}{15(k-1)(k-1)!} \cdot \ell_1 m$ for all $s \leq k-1$.*

Proof. We can derive from (4.25) that, for $s \leq k-1$ and $v \in \bigcup_{i \leq \ell_1, t \neq s} X_i^{(t)}$,

$$\begin{aligned} \deg(v, \bigcup_{i \leq \ell_1} X_i^{(s)}) &\geq \left|\bigcup_{i \leq \ell_1} X_i^{(s)}\right| + \deg(v, X^{(s)}) - \left|X^{(s)}\right| \\ &\geq \left|\bigcup_{i \leq \ell_1} X_i^{(s)}\right| - k\varepsilon n \geq \left|\bigcup_{i \leq \ell_1} X_i^{(s)}\right| - k^k \varepsilon \cdot \ell_1 m. \end{aligned}$$

Applying Lemma 4.4.6 to $G[V_1 \cup \ldots \cup V_{\ell_1}]$ with $k_{4.4.6} = k$, $\varepsilon_{4.4.6} = k^k\varepsilon$ and $\alpha_{4.4.6} = (2c)^{1/(k-1)}$, we conclude either $G[V_1 \cup \ldots \cup V_{\ell_1}]$ contains $K_{k-1}^{+}\left(\lfloor \alpha^{k-1} \log(p) \rfloor, \dots, \lfloor \alpha^{k-1} \log(p) \rfloor, \lfloor p^{1-\alpha^{k-2}} \rfloor\right)$ or there are no vertices $v \in V_1 \cup \ldots \cup V_{\ell_1}$ with $\deg(v, X_1^{(s)} \cup \ldots \cup X_\ell^{(s)}) \geq p$ for all s, where $p = \frac{1}{16(k-1)(k-1)!} \cdot \ell_1 m$. Since $\alpha^{k-1} \log(p) > c \log(n)$, $p^{1-\alpha^{k-2}} > n^{1-2\sqrt{c}}$ and since G has no copy of $K_{k-1}^{+}\big(\lfloor c\log n\rfloor, \ldots, \lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$, the former case is ruled out. The later case implies our statement. $\square$

Since $\ell_2 = \frac{\ell_1}{(k-1)!}$, by reordering parts if necessary we can assume that

$$\left|Y_i^{(1)}\right| \geq \left|Y_i^{(2)}\right| \geq \ldots \geq \left|Y_i^{(k-1)}\right| \text{ for } i \leq \ell_2. \tag{4.27}$$

For $i \leq \ell_2$ and $s \leq k-1$, let us denote

$$D_i^{(s)} = \left\{ v \in Y_i^{(s)} : \deg(v, Y_1^{(t)} \cup \ldots \cup Y_{\ell_2}^{(t)}) < \tfrac{3}{4(k-1)} \cdot \ell_2 m \text{ for some } t \neq s \right\}.$$

Claim 4.4.11. *The vertex set $\bigcup_{i\leq \ell_2} Y_i^{(s)} \setminus D_i^{(s)}$ is an independent set of G for $s \leq k-1$.*

Proof. For $i \leq \ell_2$ and $s \leq k-1$, let $B_i^{(s)}$ be the vertex set consisting of all vertices $v \in Y_i^{(s)}$ such that $\deg(v, \bigcup_{i\leq\ell_2} X_i^{(t)}) < \frac{2}{3(k-1)} \cdot \ell_2 m$ for some $t \neq s$. Note that, for $s \leq k-1$ and $v \in \bigcup_{i\leq \ell_2, t\neq s} X_i^{(s)}$, one has

$$\begin{aligned}\deg(v, \bigcup_{i\leq\ell_2} X_i^{(s)}) &\geq \left|\bigcup_{i\leq\ell_2} X_i^{(s)}\right| + \deg(v, X^{(s)}) - \left|X^{(s)}\right| \\ &\overset{(4.25)}{\geq} \left|\bigcup_{i\leq\ell_2} X_i^{(s)}\right| - k\varepsilon n \geq \left|\bigcup_{i\leq\ell_2} X_i^{(s)}\right| - k^{2k}\varepsilon \cdot \ell_2 m.\end{aligned}$$

This estimate together with Claim 4.4.9 and (4.26) show that we can apply Lemma 4.4.7 to $G[V_1 \cup \ldots \cup V_{\ell_2}]$ with $k_{4.4.7} = k$, $\varepsilon_{4.4.7} = \max\{2k^2\sqrt{\varepsilon}, k^{2k}\varepsilon\}$ and $\alpha_{4.4.7} = (2c)^{1/(k-1)} := \alpha$ to conclude that either $G[V_1 \cup \ldots \cup V_{\ell_2}]$ contains $K_{k-1}^{+}\big(\lfloor \alpha^{k-1}\log(p)\rfloor, \ldots, \lfloor \alpha^{k-1}\log(p)\rfloor, \lfloor p^{1-\alpha^{k-2}}\rfloor\big)$ or $\bigcup_{i\leq\ell_2} Y_i^{(s)} \setminus B_i^{(s)}$ is an independent set of G for $s \leq k-1$, where $p = \frac{1}{4(k-1)} \cdot \ell_2 m$. Since G has no copy of $K_{k-1}^{+}\big(\lfloor c\log n\rfloor, \ldots, \lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$ and since $\alpha^{k-1}\log(p) > c\log(n)$, $p^{1-\alpha^{k-2}} > n^{1-2\sqrt{c}}$, the former case is ruled out. We can see that the later case implies our statement. $\square$

Now we can find a large infracolourable structure in G, and then use Lemma 4.4.8 to show that G is isomorphic to a graph in $\mathcal{G}_\ell^k$.

Claim 4.4.12. *G is isomorphic to a graph in the family $\mathcal{G}_\ell^k$.*

Proof. Analogously to the proof of Claim 4.3.14, we can infer from Claims 4.4.10 and 4.4.11, (4.26) and (4.27) that $G[V_1 \cup \ldots \cup V_{\ell_2}]$ together with pairs $(D_i^{(s)}, Y_i^{(s)})_{s\leq k-1, i\leq\ell_2}$ form a $(\frac{1}{15}, k, \ell_2)$-infracolourable structure. By Lemma 4.2.3 this implies that $e(G[V_1 \cup \ldots \cup V_{\ell_2}]) \leq \binom{\ell_2}{2}\frac{k-2}{k-1}m^2$ and hence the equality must occur by the density condition. Appealing to Lemma 4.2.3

once again, we see that there exists $i_0 \in \{0, 1, \ldots, \ell_2\}$ with $\left|Y_i^{(s)}\right| = \frac{m}{k-1}$ for all s and all $i \in [\ell_2] \setminus \{i_0\}$, and $d(Y_i^{(s)}, Y_j^{(t)}) = 1$ for all $s \neq t$ and $1 \leq i < j \leq \ell_2$. Hence we can apply Lemma 4.4.8 with $\tilde{\ell} = \ell_2 - 1$ to conclude that either G contains a copy of $K_{k-1}^{+}\big(\frac{(\ell_2-1)m}{32k^2}\big)$ or G is isomorphic to a graph in $\mathcal{G}_{\ell}^{k}$. The former can not happen since G has no copy of $K_{k-1}^{+}\big(\lfloor c\log n\rfloor, \ldots, \lfloor c\log n\rfloor, \lfloor n^{1-2\sqrt{c}}\rfloor\big)$ and since $\frac{(\ell_2-1)m}{32k^2} > \max\{n^{1-2\sqrt{c}}, c\log n\}$. So G must isomorphic to a graph in the family $\mathcal{G}_{\ell}^{k}$. □

This concludes our proof of Theorem 4.3.2. □

4.5 Missing proofs

4.5.1 Proof of Theorem 4.1.5

In this section we sketch a proof of Theorem 4.1.5. We follow essentially the proof of Theorem 4.4.1. We make the following alterations. Instead of Lemma 4.4.3 we use a stability result due to Bollobás and Nikiforov [16, Theorem 9].

Lemma 4.5.1. *Let $k \geq 2$ be an integer, and let δ be a positive with $\delta < \frac{1}{16k^8}$. Suppose that G is a graph with $n > k^8$ vertices and $e(G) \geq \left(\frac{k-2}{k-1} - \delta\right)\binom{n}{2}$ edges. Then, either G contains a family of $k^{-(k+5)}n^{k-2}$ copies of K_k sharing a common edge, or G contains an induced $(k-1)$-colourable subgraph F of size $v(F) \geq (1-2\sqrt{\delta})n$ and minimum degree $\delta(F) \geq \left(\frac{k-2}{k-1} - 4\sqrt{\delta}\right)n$.*

We replace Lemma 4.4.4 by the following embedding result.

Lemma 4.5.2. *Let $r \geq 2$ be an integer, and let G be an r-colourable graph with classes $W_{(1)}, \ldots, W_{(r)}$ of the same size h. Suppose that $\deg(v, W_{(s)}) \geq \big(1 - \frac{1}{r^2}\big)h$ for $s \leq r$ and $v \in \bigcup_{t\neq s} W_{(t)}$. Then for every pair (s,t) with $s \neq t$, there is an edge between $W_{(s)}$ and $W_{(t)}$ which is contained in $\frac{1}{2}h^{r-2}$ copies of K_r.*

Proof. According to Lemma 4.4.4, G contains at least $\frac{1}{2}h^r$ copies of K_r. Hence there exists an edge between $W_{(s)}$ and $W_{(t)}$ which is shared by at least $h^r/(2h^2) = \frac{1}{2}h^{r-2}$ copies of K_r. □

The remainder of the proof is similar to that of Theorem 4.4.1.

4.5.2 Proofs of Proposition 4.3.5 and Lemma 4.3.8

To prove Proposition 4.3.5 we shall require the Erdős-Simonovits stability theorem (Erdős [34] and Simonovits [92, Theorem 8], and the graph removal lemma (Ruzsa and Szemerédi [91]).

Theorem 4.5.3 (**Stability theorem**)**.** *For every graph H and every $\varepsilon > 0$, there exist positive constants $\delta = \delta(H,\varepsilon)$ and $C = C(H,\varepsilon)$ so that the following holds for every integer $n \geq C$. Every n-vertex H-free graph with at least $\left(\frac{\chi(H)-2}{\chi(H)-1} - \delta\right)\binom{n}{2}$ edges contains a $(\chi(H)-1)$-colourable subgraph of order at least $(1-\varepsilon)n$ and minimum degree at least $\left(\frac{\chi(H)-2}{\chi(H)-1} - \varepsilon\right)n$.*

Theorem 4.5.4 (**Graph removal lemma**)**.** *For every graph H and every $\delta > 0$, there exists a positive constant $\gamma = \gamma(H,\delta)$ such that every graph on n vertices with at most $\gamma n^{v(H)}$ copies of H can be made H-free by removing from it at most $\delta\binom{n}{2}$ edges.*

Now we can deduce Proposition 4.3.5 from Theorems 4.5.3 and 4.5.4 as follows.

Proof of Proposition 4.3.5. Let $\delta = \delta_{4.5.3}(H,\varepsilon)/2$, $\gamma = \min\{\gamma_{4.5.4}(H,\delta), \delta\}$ and $C = C_{4.5.3}(H,\varepsilon)$. Since G contains at most $\gamma n^{v(H)}$ copies of H, Theorem 4.5.4 shows that G contains an H-free subgraph G' with $e(G') \geq e(G) - \delta\binom{n}{2}$. Hence

$$e(G') \geq \left(\frac{\chi(H)-2}{\chi(H)-1} - \gamma - \delta\right)\binom{n}{2} \geq \left(\frac{\chi(H)-2}{\chi(H)-1} - \delta_{4.5.3}(H,\varepsilon)\right)\binom{n}{2}.$$

Moreover, $v(G') = n \geq C = C_{4.5.3}(H,\varepsilon)$. Therefore, one can apply Theorem 4.5.3 to obtain a $(\chi(H)-1)$-colourable subgraph G'' of G' with $v(G'') \geq (1-\varepsilon)n$ and $\delta(G'') \geq \left(\frac{\chi(H)-2}{\chi(H)-1} - \varepsilon\right)n$. □

Proof of Lemma 4.3.8. Choose $D = qd^{-r}$ and $\rho = e^{-q}d^{rq}$. Let S be the set of tuples $(w_1,\dots,w_r,A)$ where $w_s \in W_{(s)}$ for all s, and $A \in \binom{N(w_1,\dots,w_r)}{q}$. We find that

$$|S| = \sum_{A\in\binom{U}{q}} \prod_{s\leq r} \left|N(A)\cap W_{(s)}\right| = \sum_{(w_1,\dots,w_r)} \binom{|N(w_1,\dots,w_r)|}{q}. \tag{4.28}$$

Moreover, our assumption implies that

$$\sum_{(w_1,\dots,w_r)} |N(w_1,\dots,w_r)| = \sum_{u\in U}\prod_{s\leq r} \deg(u, W_{(s)}) \geq d^r\,|U| \cdot \prod_{s\leq r}\left|W_{(s)}\right|. \tag{4.29}$$

Note that the function

$$\binom{x}{q} = \begin{cases} x(x-1)\cdots(x-q+1)/q! & \text{if } x \geq q-1,\\ 0 & \text{if } x < q-1.\end{cases}$$

is convex. Thus, we can first apply Jensen's inequality to the right hand side of (4.28) and then use the inequality (4.29) to obtain $|S| \geq \binom{d^r|U|}{q}\prod_{s\leq r}\left|W_{(s)}\right|$. We infer from this and the first identity in (4.28) that there is a subset $A \in \binom{U}{q}$ with

$$\prod_{s\leq r}\left|N(A)\cap W_{(s)}\right| \geq \frac{\binom{d^r|U|}{q}}{\binom{|U|}{q}} \cdot \prod_{s\leq r}\left|W_{(s)}\right| \geq e^{-q}d^{rq}\cdot\prod_{s\leq r}\left|W_{(s)}\right| = \rho\cdot\prod_{s\leq r}\left|W_{(s)}\right|,$$

where the second inequality holds since $\binom{|U|}{q} \le \left(\frac{e|U|}{q}\right)^q$, and $\binom{d^r|U|}{q} \ge \left(\frac{d^r|U|}{q}\right)^q$ for $|U| \ge D = qd^{-r} \ge q$. Hence $\left|N(A) \cap W_{(s)}\right| \ge \rho \left|W_{(s)}\right|$ for $s \le r$. $\square$

4.6 Concluding remarks

Bollobás [12, Corollary 3.5.4] showed that every n-vertex graph with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges contains cycles of lengths from 3 up to $\lfloor \frac{n+3}{2} \rfloor$, and thus strengthened the Mantel theorem. Using techniques developed in this chapter we can prove the following multipartite version of this result; we omit the details.

Theorem 4.6.1 ([81])**.** *Let $\ell \ge 10^{20}$, and let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph on n vertices such that*

$$d(V_i, V_j) \ge \tfrac{1}{2} \quad \text{for } i \ne j.$$

Then, G either contains a cycle of length h for each integer h with $3 \le h \le (\frac{1}{2} - \frac{2}{\sqrt{\ell}})n$ or is isomorphic to a graph in $\mathcal{G}_\ell^3$.

The balanced ℓ-partite graph obtained by taking the disjoint union of $K_\ell\big(\lfloor \frac{n}{2\ell} \rfloor - 1\big)$ and $K_\ell\big(\lceil \frac{n}{2\ell} \rceil + 1\big)$ has edge densities between parts strictly greater than $\frac{1}{2}$. However, every cycle of this graph has length at most $\frac{1}{2}n + 2\ell = (\frac{1}{2} + o(1))n$ provided $\ell = o(n)$. Therefore, the bound $(\frac{1}{2} - \frac{2}{\sqrt{\ell}})n$ in the above result is asymptotically best possible.

A book in a graph is a collection of triangles sharing a common edge. The size of a book is the number of triangles. Let $b(G)$ be the size of the largest book in a graph G. Generalising Mantel's theorem, Erdős [32] showed that every n-vertex graph G with $\lfloor \frac{n^2}{4} \rfloor + 1$ edges satisfies $b(G) \ge \frac{n}{6} - O(1)$. The optimal bound $b(G) \ge \lfloor \frac{n}{6} \rfloor$ was obtained independently by Edwards in an unpublished manuscript [28], and by Khadžiivanov and Nikiforov in [70]. We wonder whether a similar result holds for balanced multipartite graphs.

Conjecture 4.6.2 ([81])**.** *For every $\varepsilon > 0$, there is a constant $C = C(\varepsilon)$ such that the following holds for $\ell > C$. Let $G = (V_1 \cup \ldots \cup V_\ell, E)$ be a balanced ℓ-partite graph on n vertices such that*

$$d(V_i, V_j) > \tfrac{1}{2} \quad \text{for every } i \ne j.$$

Then, $b(G) > \left(\frac{1}{6} - \varepsilon\right) n$.

According to Theorem 4.1.5, the above conjecture is true for $\varepsilon \ge \frac{1}{6} - 3^{-18}$.

Assume H is not an almost colour-critical graph. Theorem 4.1.3(1) tells us that $d_\ell(H) \ge \frac{\chi(H)-2}{\chi(H)-1} + \frac{1}{(\chi(H)-1)^2(\ell-1)^2}$ for every $\ell \ge v(H)$. Furthermore, this estimate is tight for $H = K_{1,2}$, as shown in Remark 4.3.1. It would be very interesting to have a characterisation of the equality case.

Bondy, Shen, Thomassé and Thomassen [17] determined the value of $d_\ell(K_k)$ in the case when $\ell = k = 3$, while Pfender [85] obtained result in the case when ℓ is large enough in terms of k. The value of $d_\ell(K_k)$ is not known in the remaining cases. Nevertheless, when $\ell = k \geq 4$, Pfender [86] proposed the following conjecture (see [80, Section 5] for more details).

Conjecture 4.6.3. *The critical edge density $d_k = d_k(K_k)$ satisfies the following recurrence formula:*

$$d_2 = 0, \quad d_k^2(1 - d_{k-1}) + d_k - 1 = 0 \textit{ for } k \geq 3.$$

Finally, we emphasise that there are other interesting multipartite versions of the Turán theorem. For instance, Bollobás, Erdős and Szemerédi [14] introduced the function $\delta_r(n)$ which is the smallest integer so that every r-partite graph with parts of size n and minimum degree $\delta_r(n) + 1$ contains a copy of K_r. The exact values of $\delta_r(n)$ was determined completely by Haxell and Szabó [54] (for odd r), and Szabó and Tardos [93] (for even r) via topological methods.

Chapter 5

Keeping Avoider's graph almost acyclic

5.1 Introduction

Avoider-Enforcer games can be seen as the misère version of the well-known Maker-Breaker games (studied first by Lehman [74], Chvátal and Erdős [19] and Beck [7, 9]). This means that, while playing according to their conventional rules, the players' goal is to lose the game. The general setting of Avoider-Enforcer games can be summarized as follows. Let X be a finite set and let $\mathcal{F} \subseteq 2^X$. The two players, called *Avoider* and *Enforcer*, alternately occupy a certain number of elements of the so-called *board* X. The game ends when all elements are claimed by the players. Avoider wins if for every so-called *losing set* $F \in \mathcal{F}$, he does not occupy all elements of F by the end of the game. Otherwise Enforcer wins. In particular, it is not possible that the game ends in a draw. We may assume that Avoider is always the first player since the choice of the player who is making the first move does not have an impact on our results.

In the following we shall focus on games where the board X is given by the edge set $E(K_n)$ of a complete graph and $\mathcal{F}_n$ is some graph property to be avoided. Following Hefetz, Krivelevich, Stojaković and Szabó [56], we consider two different versions of Avoider-Enforcer games. Let b be a positive integer. In the original, *strict* $(1 : b)$ Avoider-Enforcer game (as investigated e.g. by Beck [8, 9], Hefetz, Krivelevich and Szabó [58] and by Lu [76, 77, 78]), Avoider occupies exactly 1 and Enforcer exactly b unclaimed edges per round. If the number of unclaimed edges is strictly less than b when it is Enforcer's turn, then he must select all the remaining unclaimed edges. For these strict rules, we define the lower threshold bias $f^-_{\mathcal{F}_n}$

to be the largest integer such that Enforcer has a winning strategy for the $(1 : b)$ game on $(E(K_n), \mathcal{F}_n)$ for every $b \leq f^-_{\mathcal{F}_n}$; and the upper threshold bias $f^+_{\mathcal{F}_n}$ to be the smallest non-negative integer such that Avoider has a winning strategy for every $b > f^+_{\mathcal{F}_n}$. In general, $f^-_{\mathcal{F}_n}$ and $f^+_{\mathcal{F}_n}$ do not coincide as shown by Hefetz, Krivelevich and Szabó [58].

In the *monotone* $(1 : b)$ Avoider-Enforcer game, Avoider occupies at least 1 and Enforcer at least b unclaimed edges per round. Again, if the number of unclaimed edges is strictly less than b when it is Enforcer's turn, then he must select all the remaining unclaimed edges. Games with these monotone rules are bias monotone, as it was shown by Hefetz, Krivelevich, Stojaković and Szabó in [56]. This means that there exists a unique threshold bias $f^{mon}_{\mathcal{F}_n}$ which is defined as the non-negative integer for which Enforcer wins the monotone $(1 : b)$ game if and only if $b \leq f^{mon}_{\mathcal{F}_n}$.

One might wonder at this point whether for any family $\mathcal{F}_n$ there is some general relation between the three thresholds mentioned above like $f^-_{\mathcal{F}_n} \leq f^{mon}_{\mathcal{F}_n} \leq f^+_{\mathcal{F}_n}$. Indeed, if $\mathcal{F}_n = \mathcal{F}_{P_3,n}$ is the family of all paths on 3 vertices of K_n, then these inequalities hold, as shown by Hefetz, Krivelevich, Stojaković and Szabó in [56]. However, these inequalities are not true in general and in fact the outcome of some Avoider-Enforcer games in the strict setting can differ a lot from the outcome of the corresponding monotone games. For instance, it was also shown in [56] and by Hefetz, Krivelevich and Szabó in [58] that for the Avoider-Enforcer connectivity game, where $\mathcal{F}_n = C_n$ is the family of all spanning trees of K_n, we have $f^{mon}_{C_n} = (1 + o(1))\frac{n}{\log n}$, while $f^+_{C_n} = f^-_{C_n} = \lfloor \frac{n-1}{2} \rfloor$.

In the present chapter, we shall be studying biased strict and monotone Avoider-Enforcer games, where Avoider's goal is to maintain an (almost) acyclic graph. This will have a series of improvements on the bias of various games such as planarity, colourability and minor games. Before stating our results we survey the relevant developments so far.

Define NC^k_n to be the set consisting of the edge sets of all non-k-colourable graphs on n vertices. It was proved by Hefetz, Krivelevich, Stojaković and Szabó [55] that for every $k \geq 3$, Avoider can win the strict $(1 : b)$ "non-k-colourability" game NC^k_n against any bias larger than $2kn^{1+\frac{1}{2k-3}}$. On the other hand, it was shown by the same authors [55] that there exists a constant s_k such that Enforcer has a strategy to win the game for every $b \leq s_k n$. Moreover, in the same paper the authors mention that there exists a constant $c > 0$ such that $cn \leq f^-_{NC^2_n} \leq f^+_{NC^2_n} \leq n^{3/2}$.

Let M^t_n denote the set of all edge sets of all graphs on n vertices containing a K_t-minor. Playing against a bias larger than $2n^{5/4}$, Avoider can win the strict $(1 : b)$ K_t-minor game M^t_n for every $t \geq 4$ whereas if b is almost as large as $n/2$ Enforcer has a winning strategy where

t is some constant power of n, see [55]. It was proved by Hefetz, Krivelevich, Stojaković and Szabó in [56] that the threshold bias for the monotone version is of order $n^{3/2}$ for $t = 3$.

Finally, let us introduce the "non-planarity" Avoider-Enforcer game. Let NP_n be the set consisting of the edge sets of all non-planar graphs on n vertices. In the so-called "non-planarity" game NP_n, Avoider's task is to keep his graph planar. Hefetz, Krivelevich, Stojaković and Szabó proved in [55] that in the strict $(1 : b)$ non-planarity game, Avoider can succeed against any bias larger than $2n^{5/4}$. Furthermore, their proof also can be applied when considering the monotone rules instead.

The main results of this chapter are the following two theorems. The first theorem gives a lower bound of $200n \log n$ on the bias such that both in the monotone and in the strict $(1 : b)$ Avoider-Enforcer game, Avoider can keep his graph acyclic apart from at most one unicyclic component.

Theorem 5.1.1 ([21]). *For n sufficiently large and $b \geq 200n \log n$, Avoider can ensure that both in the monotone and in the strict $(1 : b)$ Avoider-Enforcer game by the end of the game Avoider's graph is a forest plus at most one additional edge.*

In the strict $(1 : b)$ game stated in the theorem below, Avoider's task is to keep his graph acyclic for which he has again a winning strategy for some bias b between $200n \log n$ and $201n \log n$.

Theorem 5.1.2 ([21]). *For n sufficiently large, there is a bias $200n \log n \leq b \leq 201n \log n$ such that Avoider can ensure that in the strict $(1 : b)$ Avoider-Enforcer game by the end of the game Avoider's graph is a forest.*

While these results are interesting in their own right, they can be applied directly to three other games discussed above: the "non-k-colourability", the "K_t-minor", and the "non-planarity" Avoider-Enforcer games.

The two corollaries below are direct consequences of our main theorems above. In particular, these results improve upper bounds for $f^+_{NC^k_n}$ and $f^{mon}_{NC^k_n}$ with $k \geq 3$, and for $f^-_{NC^2_n}$. Furthermore better bounds are obtained for $f^+_{M^t_n}$ and $f^{mon}_{M^t_n}$ with $t \geq 4$ and for $f^-_{M^3_n}$. Finally, the bounds on $f^+_{NP_n}$ and $f^{mon}_{NP_n}$ are improved as well.

Corollary 5.1.3 ([21]). *For n sufficiently large and $b \geq 200n \log n$, Avoider can ensure that in the monotone/strict $(1 : b)$ Avoider-Enforcer game by the end of the game his graph is planar, k-colourable for $k \geq 3$, and does not contain a K_t-minor for $t \geq 4$. Thus,*

$$f^+_{NP_n}, f^+_{NC^k_n},\ f^+_{M^t_n}, f^{mon}_{NP_n}, f^{mon}_{NC^k_n},\ f^{mon}_{M^t_n} \leq 200n \log n.$$

Proof. By Theorem 5.1.1, Avoider can ensure that by the end of the game his graph is a forest plus at most one additional edge. Clearly, this graph is planar, 3-colourable, and does not contain a K_4-minor, proving the statement. $\square$

Corollary 5.1.4 ([21])**.** *For n sufficiently large, there is a bias $200n\log n \le b \le 201n\log n$ such that Avoider can ensure that in the strict $(1:b)$ Avoider-Enforcer game by the end of the game Avoider's graph is 2-colourable and does not contain a K_3-minor. Thus,*

$$f^{-}_{NC^2_n}, f^{-}_{M^3_n} = O(n\log n).$$

Proof. By Theorem 5.1.2, Avoider can ensure that by the end of the game his graph is a forest. Obviously, this graph is 2-colourable and does not contain a K_3-minor, proving the statement. $\square$

Hefetz, Krivelevich, Stojaković and Szabó conjectured in [55] that the Avoider-Enforcer non-planarity, non-k-colourability and the K_t-minor games should be *asymptotically monotone* as n tends to infinity. That is their upper and lower threshold should be of the same order, i.e. $f^{-}_{\mathcal{F}_n} = \Theta(f^{+}_{\mathcal{F}_n})$. Since in each of the three games we have lower bounds on $f^{-}_{\mathcal{F}_n}$ that are linear in n, Corollary 5.1.3 and Corollary 5.1.4 show that the threshold biases are at most $O(\log n)$ factor apart, thus giving additional evidence that this conjecture might be true.

Coming back to the $(1:b)$ non-planarity Avoider-Enforcer game, it was also proved in [55] that in the strict version Enforcer can win whenever $b \le \frac{n}{2} - o(n)$. Moreover, with a slight modification of the proof, the same result can be obtained for the monotone rules. We improve this bound as well.

Proposition 5.1.5 ([21])**.** *For n sufficiently large and $b \le 0.59n$, Enforcer can ensure that both in the monotone and in the strict $(1:b)$ Avoider-Enforcer game, Avoider creates a non-planar graph. Thus,*

$$0.59n \le f^{mon}_{NP_n},\ f^{-}_{NP_n}.$$

It should be mentioned that for the sake of readability, we do not optimize the constants in our theorems and proofs.

The rest of this chapter is organised as follows. In Section 5.2 we prove the two main results, namely Theorem 5.1.1 and Theorem 5.1.2. In Section 5.3 we study the non-planarity Avoider-Enforcer game and prove Proposition 5.1.5. Finally, in Section 5.4 we discuss some open problems.

5.2 Forests and almost forests

Proof of Theorem 5.1.1. Let n be large enough and let $b \geq 200n \log n$. In the following we shall provide Avoider with a strategy that ensures that by the end of the game Avoider's graph is a forest plus at most one additional edge.

Let t be the smallest integer with

$$n \left(\frac{t+1}{10 \log n} \right)^t < 3. \tag{5.1}$$

An easy calculation shows that $t = \Theta(\log n)$, in particular, we have for large n that

$$t < \log n/3. \tag{5.2}$$

To succeed, Avoider will play according to t stages in increasing order and each stage consists of several consecutive rounds where it is possible that a stage lasts zero rounds, i.e. that a stage does not occur at all. In the first $t-1$ stages, Avoider always claims exactly one edge in each round, connecting two components of his forest such that the sum of their sizes is minimal (whenever we talk about components, we mean the components of Avoider's forest). In the last stage, which will be shown to last at most one round, Avoider will claim an arbitrary further edge. We refer to edges, neither taken by Avoider nor by Enforcer, as *unclaimed* edges.

Starting with Stage 1, Avoider plays according to the following rules.

Stage k (for $k \in [t-1]$). If there exists an unclaimed edge e between two components T_1 and T_2 with $|V(T_1)| + |V(T_2)| = k+1$, Avoider claims such an edge, thus creating a component on the vertex set $V(T_1) \cup V(T_2)$. Then it is Enforcer's turn and the round is over.

Avoider is going to play according to Stage k in the next round as well. If there is no such edge e to be claimed at Stage k, Avoider proceeds with Stage $k+1$. (As mentioned above it might happen that there is no edge to be claimed at Stage k already when Avoider enters Stage k. In that case, this stage lasts zero rounds, and Avoider immediately proceeds with Stage $k+1$.)

Stage t. In every further round, Avoider claims exactly one arbitrary free edge.

One can easily verify that Avoider can follow the strategy. Moreover, as long as Avoider plays according to the strategy of the first $t-1$ stages, his graph remains a forest. Thus, in order to show that the above described strategy is indeed a winning strategy, it remains to show that the last stage lasts at most one round. As a first step we aim to bound the number of rounds a given stage lasts. Let n_k denote the number of rounds in Stage $k-1$. Observe that Avoider creates components of size exactly k only in this stage. Thus, the number of such components is always bounded from above by n_k.

Claim 5.2.1. *For every $k \le t$,*

$$n_k \le n \left(\frac{k}{10 \log n} \right)^{k-1}.$$

Proof. The claim is obviously true for $k = 1$. So, let $k > 1$ and we proceed by induction. When Avoider enters Stage $k-1$ every existing component contains at most $k-1$ vertices and there are no unclaimed edges between any two components T_1 and T_2 with $|V(T_1)|+|V(T_2)| \le k-1$. In particular, every unclaimed edge is either between two components T_1 and T_2 with $|V(T_1)| + |V(T_2)| \ge k$ or between two vertices within the same component which has size at most $k-1$. The first case contributes at most $\sum_{1 \le i \le j \le k-1 \colon i+j \ge k} ijn_in_j$ unclaimed edges since n_i is an upper bound on the number of components of size exactly i. For the second case we find an upper bound of $(k-1)n$ by the following reason: Let n'_i denote the number of components of order i immediately after the end of Stage $k-1$. Then the number of unclaimed edges within components after $k-1$ stages is at most $\sum_{i=1}^{k} \binom{i}{2} n'_i \le (k-1) \sum_{i=1}^{k} in'_i = (k-1)n$, since $\sum_{i=1}^{k} in'_i = n$.

Therefore, at the beginning of Stage $k-1$, the number of unclaimed edges is at most $\sum_{1 \le i \le j \le k-1 \colon i+j \ge k} ijn_in_j + (k-1)n$. Since in each but possibly the last round at least $b+1$ edges are claimed (1 by Avoider and b by Enforcer), we conclude

$$n_k \le \frac{1}{b+1} \left(\sum_{1 \le i \le j \le k-1 \colon i+j \ge k} ijn_in_j + (k-1)n \right) + 1. \tag{5.3}$$

We use the induction hypothesis to estimate the sum $\sum_{1 \le i \le j \le k-1 \colon i+j=s} ijn_in_j$ for $s = k$, $\ldots$, $2k-2$ as follows:

$$\sum_{\substack{1 \le i \le j \le k-1 \\ i+j=s}} ijn_in_j \le \frac{n^2}{(10 \log n)^{s-2}} \sum_{\substack{1 \le i \le j \le k-1 \\ i+j=s}} i^i j^j \le \frac{n^2}{(10 \log n)^{s-2}} \sum_{\substack{1 \le i \le j \le s-1 \\ i+j=s}} i^i j^j. \tag{5.4}$$

For $s \le 6$, it is easy to check that

$$\sum_{\substack{1 \le i \le j \le s-1 \\ i+j=s}} i^i j^j < 3s^{s-1}. \tag{5.5}$$

On the other hand, for $s \ge 7$ observe that we have for every $2 \le i \le s/2$

$$\left(\frac{i}{s} \right)^i \le \left(\frac{2}{s} \right)^2 \tag{5.6}$$

by an easy calculation for $i \leq 3$ and since

$$\frac{i^i}{s^{i-2}} \leq \frac{i^i}{(2i)^{i-2}} \leq \frac{i^2}{2^{i-2}} \leq 4$$

for every $i \geq 4$. Therefore, we also obtain for $s \geq 7$

$$\sum_{\substack{1\leq i\leq j\leq s-1 \\ i+j=s}} i^i j^j < s^{s-1} + \sum_{2\leq i\leq s/2} i^i s^{s-i} = s^{s-1}\left(1 + s\sum_{2\leq i\leq s/2}\left(\frac{i}{s}\right)^i\right)$$

$$\overset{(5.6)}{\leq} s^{s-1}\left(1 + s\sum_{2\leq i\leq s/2}\left(\frac{2}{s}\right)^2\right) < 3s^{s-1}. \quad (5.7)$$

Observing that

$$\left(\frac{s}{10\log n}\right)^{s-1} = \left(\frac{k}{10\log n}\right)^{k-1}\prod_{i=1}^{s-k}\frac{k+i-1}{10\log n}\left(1+\frac{1}{k+i-1}\right)^{k+i-1}$$

$$\leq \left(\frac{k}{10\log n}\right)^{k-1}\left(\frac{2ke}{10\log n}\right)^{s-k} \overset{(5.2)}{\leq} \left(\frac{k}{10\log n}\right)^{k-1} 2^{k-s}, \quad (5.8)$$

we can simplify (5.3) using $b \geq 200n\log n$ and Equations (5.4), (5.5), and (5.7) as follows

$$n_k \leq \frac{1}{200n\log n}\left(\sum_{s=k}^{2k-2} 30n^2\log n\left(\frac{s}{10\log n}\right)^{s-1} + (k-1)n\right) + 1$$

$$\overset{(5.8)}{\leq} \frac{3n}{20}\left(\frac{k}{10\log n}\right)^{k-1}\sum_{s=k}^{2k-2} 2^{k-s} + \frac{k-1}{200\log n} + 1$$

$$\overset{(5.2)}{\leq} \frac{3n}{10}\left(\frac{k}{10\log n}\right)^{k-1} + 2 \overset{(5.1)}{\leq} n\left(\frac{k}{10\log n}\right)^{k-1}.$$

This completes the proof of Claim 5.2.1. $\square$

Now, analogously to the calculation of the proof of Claim 5.2.1 it follows that, when Avoider enters the last stage, Stage t, the number of remaining unclaimed edges is bounded by

$$\sum_{\substack{1\leq i\leq j\leq t \\ i+j\geq t+1}} ijn_in_j + tn \leq \sum_{s=t+1}^{2t} 30n^2\log n\left(\frac{t+1}{10\log n}\right)^t 2^{t+1-s} + tn$$

$$\overset{(5.1)}{\leq} 180n\log n + tn < 200n\log n$$

by the choice of t ($t < \log n/3$) and for n sufficiently large. Thus, this last stage lasts at most one round. $\square$

Now we turn to the case of the strict rules, when Enforcer has to claim exactly b edges during each round (except possibly for the last one).

Proof of Theorem 5.1.2. We will show below that for large enough n, there exists b with $200n \log n \leq b \leq 201n \log n$ and the remainder of $\binom{n}{2}$ divided by $b+1$ is at least $n \log n$.

Before proving this claim let us explain how the theorem follows then. Let b be given as above. Avoider now plays according to the same strategy as given in the proof of Theorem 5.1.1 until he reaches Stage t, where again t is the smallest integer with $n\left(\frac{t+1}{10\log n}\right)^t < 3$. At this point, Avoider's graph is still a forest, the components of which are all of size at most t. Analogously to the proof of Theorem5.1.1, there can be at most $tn < n\ln n/3$ unclaimed edges within components. However, since the remainder of the division $\binom{n}{2}/(b+1)$ is at least $n\log n$, there exist unclaimed edges connecting two different components when Avoider enters Stage t (provided n is large enough). Now, Avoider just claims one such edge arbitrarily. His graph remains a forest and afterwards, Enforcer must take all remaining edges. Observe that in the case when Avoider is the second player, he does not even claim an edge in the last round.

So, it only remains to prove the above mentioned claim. Let $b_1 = \lceil 200.5n\log n\rceil$. Moreover, let

$$\binom{n}{2} = q_1(b_1+1) + r_1 \quad \text{with} \quad 0 \leq r_1 \leq b_1 \text{ and } q_1 \sim \frac{n}{401\log n}.$$

If $r_1 > n\ln n$, we are done by setting $b = b_1$. Otherwise, let $b = b_1 - \lceil 402\log^2 n\rceil$. Then

$$\binom{n}{2} = q_1(b+1) + (r_1 + q_1\lceil 402\log^2 n\rceil).$$

Moreover, for large enough n, we obtain $r_1 + q_1\lceil 402\log^2 n\rceil < b$, and therefore the remainder of the division $\binom{n}{2}$ by $(b+1)$ is at least $r_1 + q_1\lceil 402\log^2 n\rceil > n\log n$, while $200n\log n \leq b \leq 201n\log n$. $\square$

5.3 Lower bound in the non-planarity game

Before obtaining a lower bound for the non-planarity Avoider-Enforcer game in Proposition 5.1.5, we analyze another strict game where two players, the first player (denoted by FP) and the second player (denoted by SP), claim *exactly* 1 and b edges, respectively.

Proposition 5.3.1. *Let $c = 1/1000$. For n sufficiently large and every $0.49n \leq b \leq 0.59n$ the second player in a strict $(1:b)$ game on $E(K_n)$ can isolate at least*

$$n - (1-c)\frac{n^2}{2b} \qquad \textit{vertices,}$$

i.e. claim all edges that are incident to these vertices.

Proof. **Case 1. ($\mathbf{0.49n \le b \le 0.55n}$.)** As long as there are at least 4 vertices not isolated by the second player (SP) and not touched by the first player (FP), SP can isolate a vertex in every fourth round. Indeed, assume SP isolated a vertex in the previous round and now wants to isolate one vertex within the next 4 rounds. He fixes 4 vertices v_1, v_2, v_3, v_4 that are neither isolated by him nor touched by FP. In every first round, SP claims all edges between these 4 vertices and at each v_i he additionally claims $\lfloor (b-6)/4 \rfloor$ arbitrary incident edges. Now, it is FP's turn. He can touch at most one of these four vertices since all edges between them are already claimed by SP. Without loss of generality, v_1, v_2, and v_3 are still untouched by FP. Now in the second round SP claims at each of these three vertices $\lfloor b/3 \rfloor$ arbitrary incident edges. Again, FP can touch at most one of these three vertices in his turn. Without loss of generality, v_1 and v_2 are still untouched by FP after that. In the third round, SP claims at each of these two vertices $\lfloor b/2 \rfloor$ arbitrary incident edges. After FP's next turn, w. l. o. g. v_1 is still untouched by FP. Now, SP simply claims all remaining incident edges at v_1, which is possible since $3 + \lfloor (b-6)/4 \rfloor + \lfloor b/3 \rfloor + \lfloor b/2 \rfloor + b > n$, for large n. Note that while SP isolates one vertex, FP can touch at most 8 other vertices. It follows that the number of vertices that SP isolates in total is at least $\lfloor n/9 \rfloor \ge n - (1-c)\frac{n^2}{2b}$.

Case 2. ($\mathbf{0.55n \le b \le 0.58n}$.) Analogously to Case 1, SP can isolate a vertex in every third round as long as there are at least 3 vertices not touched by FP. This time, SP starts by only fixing three vertices v_1, v_2, v_3 and isolates then one of them within three rounds, which is possible since $2 + \lfloor (b-3)/3 \rfloor + \lfloor b/2 \rfloor + b > n$, for large n. It follows then that SP isolates at least $\lfloor n/7 \rfloor \ge n - (1-c)\frac{n^2}{2b}$ vertices in total.

Case 3. ($\mathbf{0.58n \le b \le 0.59n}$.) Analogously to Case 2, SP can isolate a vertex in every third round as long as there are at least 3 vertices not touched by FP. In a first phase, SP follows the above described strategy and he isolates $n - 1.5b$ vertices, which happens in at most $3n - 4.5b$ rounds. During this phase, FP can touch at most $6n - 9b$ vertices. Afterwards, for every vertex that is neither isolated by SP nor touched by FP, SP only needs to claim at most $1.5b$ further incident edges in order to isolate it. But then, analogously to the previous cases, SP can isolate one vertex in every second round, since $1 + \lfloor (b-1)/2 \rfloor + b \ge 1.5b$. Thus, in the second phase after at most $3n - 4.5b$ rounds, SP isolates a vertex in every second round as long as possible. Since at the beginning of the second phase at least $n - (n - 1.5b) - (6n - 9b) = 10.5b - 6n$ vertices were neither isolated by SP nor touched by FP, SP can isolate at least $(10.5b - 6n)/5$ further vertices. In total SP will isolate at least $(n - 1.5b) + (10.5b - 6n)/5 \ge n - (1-c)\frac{n^2}{2b}$ vertices. □

Lemma 5.3.2. *For n sufficiently large and $b \le 0.59n$ Enforcer can ensure that in the strict $(1:b)$ game on $E(K_n)$ Avoider creates a non-planar graph. Thus,*

$$0.59n \le f^{-}_{NP_n}.$$

Proof. Since the statement is already proved for $b \le 0.49n$ in [55], we may assume from now on that $0.49n \le b$. The following proof will be a slight modification of the one given in [55]. Let $c = 1/1000$ be as in Proposition 5.3.1 and choose an integer $k \ge 3$ such that

$$\frac{k}{k-2}\left(1-\frac{c}{2}\right) < 1. \tag{5.9}$$

Enforcer's strategy consists of two goals: First of all, he wants to prevent Avoider from creating cycles of length at most k. Secondly, he wants to isolate a large number of vertices to ensure that Avoider's graph lives on a small vertex set. For this he splits his bias $b = b_1 + b_2$ (b_1 and b_2 will be chosen later) and uses b_1 for his first goal, and b_2 for the second goal.

Preventing cycles. It follows from the work of Bednarska and Łuczak [10] (see also the proof of Theorem 2.3 in [55]), that for every $3 \le i \le k$ there is a constant c_i such that, for sufficiently large n, Enforcer can prevent Avoider from claiming a cycle of length i if Enforcer is allowed to claim at least $c_i n^{\frac{i-2}{i-1}}$ edges. Let $C = \max\{c_i : \ 3 \le i \le k\}$. Then, simultaneously playing according to the different strategies for preventing cycles of length $3 \le i \le k$, Enforcer can ensure that Avoider's graph has girth larger than k if he claims at least

$$\sum_{i=3}^{k} c_i n^{\frac{i-2}{i-1}} \le Ckn^{\frac{k-2}{k-1}} =: b_1$$

edges per round. Observe that $b_1 = o(b)$.

Isolating vertices. Let $b_2 = b - b_1 = b(1-o(1))$. In each round Enforcer uses b_2 edges to play according to the strategy given in the proof of Proposition 5.3.1. Therefore, he isolates at least $n - (1-c)\frac{n^2}{2b_2} \ge n - \left(1-\frac{c}{2}\right)\frac{n^2}{2b}$ vertices.

Now, let Enforcer split his bias $b = b_1 + b_2$, and thus play so as to prevent cycles of length at most k, while at the same time to isolate at least $n - \left(1-\frac{c}{2}\right)\frac{n^2}{2b}$ vertices. Notice that it does not hurt Enforcer if the combination of the above strategies leads to claiming the same edge more than once - Enforcer can claim an arbitrary edge instead since this does not destroy the properties of the graph he is about to create. Let A be Avoider's graph at the end of the game. We know that $|V(A)| \le \left(1-\frac{c}{2}\right)\frac{n^2}{2b}$ and $\text{girth}(A) > k$. If A was planar, then, by a standard application of Euler's formula, we would have

$$|E(A)| < \frac{k}{k-2}\left(|V(A)|-2\right) < \frac{k}{k-2}\left(1-\frac{c}{2}\right)\frac{n^2}{2b}.$$

However, by the number of rounds the game lasts, we have

$$|E(A)| \geq \left\lfloor \frac{\binom{n}{2}}{b+1} \right\rfloor > \frac{k}{k-2}\left(1-\frac{c}{2}\right)\frac{n^2}{2b},$$

using (5.9), for n sufficiently large. Thus, Avoider's graph is non-planar and Enforcer wins. $\square$

Lemma 5.3.3. *For n sufficiently large and $b \leq 0.59n$ Enforcer can ensure that Avoider creates a non-planar graph in the monotone $(1:b)$ game on $E(K_n)$. Thus,*

$$0.59n \leq f^{mon}_{NP_n}.$$

Proof. Let A be Avoider's graph throughout the game, and let $A^* \subseteq A$ be a subgraph consisting of exactly one edge from every round played so far. Enforcer claims in every round exactly $b' := \max\{0.49n, b\}$ edges according to the strategy given in the proof of the previous lemma, assuming A^* to be Avoider's graph. If this strategy asks Enforcer to claim an edge from $A \setminus A^*$, he will claim another arbitrary edge instead. We distinguish two cases.

Case 1. $|\mathbf{E}(\mathbf{A})| > \mathbf{3n}$. Then, by Euler's formula, Avoider's graph is non-planar and Enforcer wins.

Case 2. $|\mathbf{E}(\mathbf{A})| \leq \mathbf{3n}$. Then the number of rounds the game lasts is at least $\frac{\binom{n}{2}-3n}{b'} = \frac{n^2}{2b'}(1-o(1))$, which also gives

$$|E(A^*)| \geq \frac{n^2}{2b'}(1-o(1)).$$

By the above described strategy we get again, similar to the proof of Lemma 5.3.2, $|V(A^*)| \leq \left(1-\frac{c}{2}\right)\frac{n^2}{2b'}$ as well as $\mathrm{girth}(A^*) > k$, ensuring that A^* cannot be planar provided that n is large enough. $\square$

Proof of Proposition 5.1.5. This proposition follows directly from Lemma 5.3.2 and Lemma 5.3.3. $\square$

5.4 Concluding remarks

For each of the games considered for Corollary 5.1.3, we have shown that the lower and upper threshold bias differ at most by a factor of $\ln n$. However, we believe that this factor can be replaced by some constant. We wonder whether this can already be done for the strategy we analyzed in the proof of Theorem 5.1.1, where we have shown that Avoider can keep his graph almost acyclic.

Question 5.4.1 ([21])**.** *Is there a constant $C > 0$ such that the following holds: For n sufficiently large and $b \geq Cn$, Avoider has a strategy that creates at most one cycle in the monotone/strict $(1 : b)$ game?*

In case the question above can be answered positively, the following conjecture by Hefetz, Krivelevich, Stojaković and Szabó [55, Conjecture 5.2] would follow immediately.

Conjecture 5.4.2. *The Avoider-Enforcer non-planarity, non-k-colourability and K_t-minor games are asymptotically monotone for every $k \geq 3$ and $t \geq 4$.*

Our result on the lower threshold bias for the non-planarity game is obtained by splitting Enforcer's strategy into two parts. The first part, based on the strategy from [55], is to prevent small cycles in Avoider's graph. The second part is to isolate a large number of vertices. So, our improvement was obtained by studying a positional game in which one player has the goal to isolate as many vertices as possible. This game itself seems to be of interest.

Question 5.4.3 ([21])**.** *Let $b \in \mathbb{N}$. What is the largest number of vertices that the second player can isolate in a $(1 : b)$ game on $E(K_n)$ under the strict rules?*

Zusammenfassung

Das Gebiet der Extremalen Kombinatorik beschäftigt sich mit der folgenden grundlegenden Frage: ,,Wie groß kann eine Struktur sein, wenn sie keine verbotenen Teilstrukturen enthält?" Die studierten Strukturen sind dabei äußerst flexibel, sodass sie eine große Bandbreite an Anwendungen in verschiedensten Fachgebieten, wie der Theoretischen Informatik, dem Operations Research, der Diskreten Geometrie und der Zahlentheorie, ermöglichen. Vieles in der Extremalen Kombinatorik betrifft Klassen von Mengen, was Extremale Mengentheorie genannt wird. Zum Beispiel: Was ist die größte Anzahl k-elementiger Teilmengen einer n-elementigen Menge, die sich paarweise schneiden können? Die Antwort auf diese Frage, nämlich der Satz von Erdős-Ko-Rado, hatte viele Fragestellungen in der Extremalen Mengenlehre zu Folge.

In dieser Dissertation stellen wir neue Erweiterungen klassischer Theoreme der Extremalen Kombinatorik vor, wobei wir probabilistische und analytische Argumente verwenden. In Kapitel 2 nutzen wir die analytische Methode, um die Stabilität des Erdős-Ko-Rado-Theorems zu untersuchen. Insbesondere zeigen wir, dass jedes Mengensystem, welches wenige disjunkte Paare enthält, durch Wegnahme weniger Mengen überschneidend gemacht werden kann. Mit diesem Resultat ausgerüstet, klären wir eine Frage von Bollobás, Narayanan and Raigorodski (2014) bezüglich der Unabhängigkeitszahl von Zufallsgraphen des Knesergraphs. In Kapitel 3 untersuchen wir eine Variante des ursprünglichen Problems von Erdős und Rothschild, in dem wir disjunkte Paare von Hyperkanten gleicher Farbe verbieten. Unsere Resultate erweitern das Erdős-Ko-Rado-Theorem. Danach schreiten wir zur Extremalen Graphentheorie und studieren in Kapitel 4 eine multipartite Version des Satzes von Turán, wobei wir die probabilistische Methode in Verbindung mit einem Stabilitätsansatz verwenden.

Der letzte Teil dieser Arbeit behandelt Avoider-Enforcer-Spiele, in denen zwei Spieler abwechselnd Kanten des vollständigen Graphs einnehmen. Avoider gewinnt, wenn es ihr gelingt zu vermeiden, dass sie alle Kanten einer Verlierermenge besetzt. In diesem Kapitel erhalten wir im Wesentlichen optimale obere Schranken für die ,,threshold biases" des ,,non-planarity"-Spiels und des ,,non-k-colourability"-Spiels, womit wir eine Frage von Hefetz, Krivelevich, Stojaković und Szabó (2008) aufgreifen und deren Resultate bedeutend verbessern.

Eidesstattliche Erklärung

Gemäß §7 (4) der Promotionsordnung des Fachbereichs Mathematik und Informatik der Freien Universität Berlin versichere ich hiermit, dass ich alle Hilfsmittel und Hilfen angegeben und auf dieser Grundlage die Arbeit selbständig verfasst habe. Des Weiteren versichere ich, dass ich diese Arbeit nicht schon einmal zu einem früheren Promotionsverfahren eingereicht habe.

Berlin, den

Manh Tuan Tran

For reasons of data protection, the curriculum vitae is not published in the electronic version.

Bibliography

[1] R. Ahlswede and L. H. Khachatrian, *The complete nontrivial-intersection theorem for systems of finite sets*, Journal of Combinatorial Theory, Series A **76** (1996), 121–138.

[2] N. Alon, J. Balogh, P. Keevash and B. Sudakov, *The number of edge colorings with no monochromatic cliques*, Journal of London Mathematical Society **70** (2004), 273–288.

[3] J. Balogh, T. Bohman and D. Mubayi, *Erdős-Ko-Rado in random hypergraphs*, Combinatorics, Probability and Computing **18** (2009), 629–646.

[4] J. Balogh, B. Bollobás and B. Narayanan, *Transference for the Erdős-Ko-Rado theorem*, Forum of Mathematics, Sigma **3** (2015).

[5] J. Balogh, S. Das, M. Delcourt, H. Liu and M. Sharifzadeh, *Intersecting families of discrete structures are typically trivial*, Journal of Combinatorial Theory, Series A **132** (2015), 224–245.

[6] Z. Baranyai, *On the factorization of the complete uniform hypergraph*, Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday), vol. 1, 1975, pp. 91–108.

[7] J. Beck, *Remarks on positional games. I*, Acta Mathematica Hungarica **40** (1982), 65–71.

[8] ———, *Ramsey games*, Discrete Mathematics **249** (2002), 3–30.

[9] ———, **Combinatorial Games: Tic-Tac-Toe Theory**, Cambridge University Press, New York, 2008.

[10] M. Bednarska and T. Łuczak, *Biased positional games for which random strategies are nearly optimal*, Combinatorica **20** (2000), 477–488.

[11] A. Blokhuis, A. E. Brouwer, A. Chowdhury, P. Frankl, T. Mussche, B. Patkós and T. Szőnyi, *A Hilton-Milner theorem for vector spaces*, The Electronic Journal of Combinatorics **17** (2010), R71.

[12] B. Bollobás, **Extremal graph theory**, London Mathematical Society Monographs, vol. 11, London, 1978.

[13] ______, **Modern graph theory**, vol. 184, Springer-Verlag, New York, 1998.

[14] B. Bollobás, P. Erdős and E. Szemerédi. *On complete subgraphs of r-chromatic graphs*, Discrete Mathematics **13** (1975), 97–107.

[15] B. Bollobás, B. Narayanan and A. Raigorodskii, *On the stability of the Erdős-Ko-Rado theorem*, Journal of Combinatorial Theory, Series A, **137** (2016), 64–78.

[16] B. Bollobás and V. Nikiforov, *Joints in graphs*, Discrete Mathematics **308** (2008), 9–19.

[17] A. Bondy, J. Shen, S. Thomassé and C. Thomassen, *Density conditions for triangles in multipartite graphs*, Combinatorica **26** (2006), 121–131.

[18] A. Chowdhury and B. Patkós, *Shadows and intersections in vector spaces*, Journal of Combinatorial Theory, Series A **117** (2010), 1095–1106.

[19] V. Chvátal and P. Erdős, *Biased positional games*, Annals of Discrete Mathematics **2** (1978), 221–229.

[20] D. Clemens, S. Das and T. Tran, *Erdős-Rothschild problem for intersecting families*, in preparation.

[21] D. Clemens, J. Ehrenmüller, Y. Person and T. Tran, *Keeping Avoider's graph almost acyclic*, The Electronic Journal of Combinatorics **22** (2015), P1.60.

[22] D. Conlon and J. Fox, *Graph removal lemmas*, Surveys in combinatorics **409** (2013), 1–50.

[23] S. Das, W. Gan and B. Sudakov, *The minimum number of disjoint pairs in set systems and related problems*, Combinatorica (2016), to appear.

[24] S. Das and B. Sudakov, *Most probably intersecting hypergraphs*, The Electronic Journal of Combinatorics **22** (2015), P1.80.

[25] S. Das and T. Tran, *Removal and stability for Erdős-Ko-Rado*, SIAM Journal on Discrete Mathematics **30** (2016), 1102–1114.

[26] P. Devlin and J. Kahn, *On "stability" in the Erdős-Ko-Rado theorem*, SIAM Journal on Discrete Mathematics **30** (2016), 1283–1289.

[27] I. Dinur and E. Friedgut, *Intersecting families are essentially contained in juntas*, Combinatorics, Probability and Computing **18** (2009), 107–122.

[28] C. Edwards, *A lower bound for the largest number of triangles with a common edge*, Unpublished manuscript (1977).

[29] D. Ellis, *Stability for t-intersecting families of permutations*, Journal of Combinatorial Theory, Series A **118** (2011), 208–227.

[30] ———, *personal communication.*

[31] D. Ellis, E. Friedgut and H. Pilpel, *Intersecting families of permutations*, Journal of the American Mathematical Society **24** (2011), 649–682.

[32] P. Erdős, *On a theorem of Rademacher-Turán*, Illinois Journal of Mathematics **6** (1962), 122–127.

[33] ———, *On the structure of linear graphs*, Israel Journal of Mathematics **1** (1963), 156–160.

[34] ———, *Some recent results on extremal problems in graph theory*, Results, Theory of Graphs (Internat. Sympos., Rome, 1966), Gordon and Breach, New York (1967), 117–123.

[35] ———, *On the number of complete subgraphs and circuits contained in graphs*, Časopis pro pěstování matematiky **94** (1969), 290–296.

[36] ———, *Some new applications of probability methods to combinatorial analysis and graph theory*, University of Calgary, Department of Mathematics, Statistics and Computing Science, 1974.

[37] P. Erdős, C. Ko and R. Rado, *Intersection theorems for systems of finite sets*, The Quarterly Journal of Mathematics **12** (1961), 313–320.

[38] P. Erdős and M. Simonovits, *A limit theorem in graph theory*, Studia Scientiarum Mathematicarum Hungarica, 1966.

[39] P. Erdős and V. T. Sós, *On Ramsey-Turán type theorems for hypergraphs*, Combinatorica **2** (1982), 289–295.

[40] P. Erdős and A. Stone, *On the structure of linear graphs*, Bulletin of the American Mathematical Society **52** (1946), 1087–1091.

[41] Y. Filmus, *Friedgut-Kalai-Naor theorem for slices of the boolean cube*, submitted, available online at http://arxiv.org/abs/1410.7834

[42] P. Frankl, *The Erdős-Ko-Rado theorem is true for $n = ckt$*, Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), vol. 1, 1978, pp. 365–375.

[43] ———, *Erdős-Ko-Rado theorem with conditions on the maximal degree*, Journal of Combinatorial Theory, Series A **46** (1987), 252–263.

[44] P. Frankl and M. Deza, *On the maximum number of permutations with given maximal or minimal distance*, Journal of Combinatorial Theory, Series A **22** (1977), 352–360.

[45] P. Frankl and N. Tokushige, *Some inequalities concerning cross-intersecting families*, Combinatorics, Probability and Computing **7** (1998), 247–260.

[46] P. Frankl and R. M. Wilson, *The Erdös-Ko-Rado theorem for vector spaces*, Journal of Combinatorial Theory, Series A **43** (1986), 228–236.

[47] E. Friedgut, *On the measure of intersecting families, uniqueness and stability*, Combinatorica **28** (2008), 503–528.

[48] E. Friedgut, G. Kalai and A. Naor, *Boolean functions whose Fourier transform is concentrated on the first two levels*, Advances in Applied Mathematics **29** (2002), 427–437.

[49] E. Friedgut and O. Regev, personal communication.

[50] Z. Füredi, *Cross-intersecting families of finite sets*, Journal of Combinatorial Theory, Series A **72** (1995), 332–339.

[51] M. Gauy, H. Hàn and I. Oliveira, *Erdős-Ko-Rado for random hypergraphs: asymptotics and stability*, Combinatorics Probability and Computing (2016), to appear.

[52] A. Hamm and J. Kahn, *On Erdős-Ko-Rado for random hypergraphs I*, submitted, available online at arXiv:1412.5085.

[53] ———, *On Erdős-Ko-Rado for random hypergraphs II*, submitted, available online at arXiv:1406.5793.

[54] P. Haxell and T. Szabó, *Odd independent transversals are odd*, Combinatorics, Probability and Computing **15** (2006), 193–211.

[55] D. Hefetz, M. Krivelevich, M. Stojaković and T. Szabó, *Planarity, colorability, and minor games*, SIAM Journal on Discrete Mathematics **22** (2008), 194–212.

[56] ———, *Avoider-Enforcer: the rules of the game*, Journal of Combinatorial Theory, Series A **117** (2010), 152–163.

[57] ———, **Positional Games (Oberwolfach Seminars)**, Birkhäuser, 2014.

[58] D. Hefetz, M. Krivelevich and T. Szabó, *Avoider-Enforcer games*, Journal of Combinatorial Theory, Series A **114** (2007), 840–853.

[59] A. Hilton and E. Milner, *Some intersection theorems for systems of finite sets*, The Quarterly Journal of Mathematics **18** (1967), 369–384.

[60] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, Journal of the American Statistical Association **58** (1963), 13–30.

[61] C. Hoppen, Y. Kohayakawa and H. Lefmann, *Edge colourings of graphs avoiding monochromatic matchings of a given size*, Combinatorics, Probability and Computing **21** (2012), 203–218.

[62] ———, *Hypergraphs with many Kneser colorings*, European Journal of Combinatorics **33** (2012), 816–843.

[63] ———, *Edge-colorings of graphs avoiding fixed monochromatic subgraphs with linear Turán number*, European Journal of Combinatorics **35** (2014), 354–373.

[64] ———, *Edge-colorings of uniform hypergraphs avoiding monochromatic matchings*, Discrete Mathematics **338** (2015), 262–271.

[65] C. Hoppen, H. Lefmann and K. Odermann, *A coloring problem for intersecting vector spaces*, Discrete Mathematics **339** (2016), 2941–2954.

[66] W. N. Hsieh, *Intersection theorems for systems of finite vector spaces*, Discrete Mathematics **12** (1975), 1–16.

[67] G. O. H. Katona, G. Y. Katona and Z. Katona, *Most probably intersecting families of subsets*, Combinatorics, Probability and Computing **21** (2012), 219–227.

[68] P. Keevash, *Shadows and intersections: stability and new proofs*, Advances in Mathematics **218** (2008), 1685–1703.

[69] P. Keevash and D. Mubayi, *Set systems without a simplex or a cluster*, Combinatorica **30** (2010), 175–200.

[70] N Khadžiivanov and V. Nikiforov, *Solution of a problem of P. Erdős about the maximum number of triangles with a common edge in a graph*, Comptes Rendus De L'Académie Bulgare Des Sciences **32** (1979), 1315–1318.

[71] H. Lefmann and Y. Person, *Exact results on the number of restricted edge colorings for some families of linear hypergraphs*, Journal of Graph Theory **73** (2013), 1–31.

[72] H. Lefmann, Y. Person, V. Rödl and M. Schacht, *On colourings of hypergraphs without monochromatic Fano planes*, Combinatorics, Probability and Computing **18** (2009), 803–818.

[73] H. Lefmann, Y. Person and M. Schacht, *A structural result for hypergraphs with many restricted edge colorings*, Journal of Combinatorics **1** (2010), 441–475.

[74] A. Lehman, *A solution of the Shannon switching game*, Journal of the Society for Industrial and Applied Mathematics **12** (1964), 687–725.

[75] L. Lovász, *On the Shannon capacity of a graph*, IEEE Transactions on Information Theory **25** (1979), 1–7.

[76] X. Lu, *A matching game*, Discrete Mathematics **94** (1991), 199–207.

[77] ———, *Hamiltonian games*, Journal of Combinatorial Theory, Series B **55** (1992), 18–32.

[78] ———, *A Hamiltonian game on $K_{n,n}$*, Discrete Mathematics **142** (1995), 185–191.

[79] W. Mantel, *Problem 28*, Wiskundige Opgaven **10** (1907), 60–61, 320.

[80] Z. Nagy, *A multipartite version of the Turán problem - density conditions and eigenvalues*, The Electronic Journal of Combinatorics **18** (2011), P46.

[81] L. Narins and T. Tran, *A Density Turán Theorem*, Journal of Graph Theory, to appear. DOI: 10.1002/jgt.22075

[82] V. Nikiforov, *Turán's theorem inverted*, Discrete Mathematics **310** (2010), 125–131.

[83] ———, *Some new results in extremal graph theory*, Surveys in Combinatorics (ed. R. Chapman), LMS Lecture Note Series **392** (2011), 141–183.

[84] O. Pikhurko, K. Staden and Z. Yilma, *The Erdős-Rothschild problem on edge-colourings with forbidden monochromatic cliques*, submitted, available online at https://arxiv.org/abs/1605.05074

[85] F. Pfender, *Complete subgraphs in multipartite graphs*, Combinatorica **32** (2012), 483–495.

[86] ———, personal communication.

[87] O. Pikhurko and Z. Yilma, *The Maximum Number of K_3-Free and K_4-Free Edge 4-Colorings*, Journal of the London Mathematical Society, **85** (2012), 593–615.

[88] E. Sperner, *Ein Satz über Untermengen einer endlichen Menge*, Mathematische Zeitschrift **27** (1928), 544–548.

[89] P. Russell, *Compressions and probably intersecting families*, Combinatorics, Probability and Computing **21** (2012), 301–313.

[90] P. Russell and M. Walters, *Probably intersecting families are not nested*, Combinatorics, Probability and Computing **22** (2013), 146–160.

[91] I. Ruzsa and E. Szemerédi, *Triple systems with no six points carrying three triangles*, Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai **18** (1978), 939–945.

[92] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), 1968, pp. 279–319.

[93] T. Szabó and G. Tardos, *Extremal problems for transversals in graphs with bounded degree*, Combinatorica **26** (2006), 333–351.

[94] P. Turán, *On an extremal problem in graph theory*, Matematikai és Fizikai Lapok **48** (1941), 436–452.

[95] R. M. Wilson, *The exact bound in the Erdős-Ko-Rado theorem*, Combinatorica **4** (1984), 247–257.

[96] R. Yuster, *The number of edge colorings with no monochromatic triangle*, Journal of Graph Theory **21** (1996), 441–452.