VIEWING EXTREMAL AND STRUCTURAL PROBLEMS THROUGH A PROBABILISTIC LENS

BY

MICHELLE DELCOURT

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral Committee:

Professor Alexandr V. Kostochka, Chair
Professor József Balogh, Director of Research
Associate Professor Kay Kirkpatrick
Assistant Professor Anush Tserunyan

Abstract

This thesis focuses on using techniques from probability to solve problems from extremal and structural combinatorics.

The main topic in Chapter 2 is determining the typical structure of t-intersecting families in various settings and enumerating such systems. The analogous sparse random versions of our extremal results are also obtained. The proofs follow the same general framework, in each case using a version of the Bollobás Set-Pairs Inequality to bound the number of maximal intersecting families, which then can be combined with known stability theorems. Following this framework from joint work with Balogh, Das, Liu, and Sharifzadeh, similar results for permutations, uniform hypergraphs, and vector spaces are obtained.

In 2006, Barát and Thomassen conjectured that the edges of every planar 4-edge-connected 4-regular graph can be decomposed into disjoint copies of S_3, the star with three leaves. Shortly afterward, Lai constructed a counterexample to this conjecture. Following joint work with Postle, in Chapter 3 using the Small Subgraph Conditioning Method of Robinson and Wormald, we find that a random 4-regular graph has an S_3-decomposition asymptotically almost surely, provided we have the obvious necessary divisibility conditions.

In 1988, Thomassen showed that if G is at least $(2k-1)$-edge-connected then G has a spanning, bipartite k-connected subgraph. In 1989, Thomassen asked whether a similar phenomenon holds for vertex-connectivity; more precisely: is there an integer-valued function $f(k)$ such that every $f(k)$-connected graph admits a spanning, bipartite k-connected subgraph? In Chapter 4, as in joint work with Ferber, we show that every $10^{10}k^3 \log n$-connected graph admits a spanning, bipartite k-connected subgraph.

Dedicated to my parents.

Acknowledgments

First, I would like to thank Józsi for all of his support and guidance. I learned a considerable amount both mathematically and professionally from my time with him and am truly fortunate to have been his student.

A number of professors and mentors from my undergraduate studies inspired and encouraged me to pursue mathematics; I especially would like to mention Neil Calkin, Ernie Croot, Michael Loss, Enid Steinbart, Prasad Tetali, Robin Thomas, Tom Trotter, and Xingxing Yu as significant figures in my life. During my time in graduate school, I had the pleasure to interact with a number of talented and passionate individuals. My professors, in particular Jayadev Athreya, Józsi Balogh, Zoltan Füredi, Kay Kirkpatrick, Sasha Kostochka, Anush Tserunyan, and Doug West, all have such enthusiasm for mathematics both inside and outside of the classroom, and I have enjoyed many interesting conversations with all of them. I also would like to thank Béla Csaba, Péter Hajnal, and Andás Pluhár for their hospitality and fruitful discussions during my visit to the University of Szeged in Hungary.

My friends and colleagues have helped to foster a stimulating and supportive environment. I have particularly enjoyed collaborating with Anton Bernshteyn, Shagnik Das, Asaf Ferber, Bernard Lidický, Hong Liu, Cory Palmer, and Maryam Sharifzadeh. For all of their help and encouragement, I also would like to thank Ruth Davidson and Rosemary Guzman. A number of fellow graduate students including Jane Butterfield, Sarah Loeb, Sogol Jahanbekam, Šárka Petříčková, Adam Wagner, Kelly Yancey, and Matt Yancey have made my years at Illinois enjoyable.

The encouragement of my family has meant so much to me; my parents and sister have supported me every step of the way. I would also like to thank Uncle Dan and Uncle Eric for sending me books and letters and encouraging my interest in solving problems and puzzles from a young age. I would like to thank Hannah Spinoza for welcoming me into her family; she, Magnolia, and Julian have made my time in grad school more entertaining. Finally, I would like to thank the love of my life, Luke Postle, for all of his encouragement and advice; I look forward to many years of fun adventures and beautiful theorems together.

I would also like to acknowledge NSF Graduate Research Fellowship DGE 1144245.

Table of Contents

List of Figures

Symbols and Notation

$\emptyset$	the empty set
$\simeq$	isomorphic as graphs
$\gg$	much greater
$\ll$	much less
$[n]$	for $n \in \mathbb{N}$, $[n] := \{1, 2, \ldots, n\}$
$\alpha(G)$	the independence number of (hyper)graph G (the size of a largest independent set)
$\chi(G)$	the chromatic number of (hyper)graph G
$E(G)$	the edge set of (hyper)graph G
$e(G)$	$e(G) := \lvert E(G) \rvert$ the size of the edge set of (hyper)graph G
$I(n, k)$	the number of intersecting k-uniform hypergraphs on $[n]$
$\kappa(G)$	the vertex-connectivity of G
log	logarithm base 2
ln	logarithm base e
$\mathbb{N}$	the set of natural numbers (1, 2, 3, ...)
$\mathbb{R}$	the set of real numbers
S_n	the symmetric group on $[n]$
$V(G)$	the vertex set of (hyper)graph G
$v(G)$	$v(G) := \lvert V(G) \rvert$ the size of the vertex set of (hyper)graph G

Chapter 1

Introduction

One will not get anywhere in graph theory by sitting in an armchair and trying to ***understand graphs*** *better. Neither is it particularly necessary to read much of the literature before tackling a problem: it is of course helpful to be aware of some of the most important techniques, but the interesting problems tend to be open precisely because the established techniques cannot easily be applied.*

– W. T. Gowers, *The Two Cultures of Mathematics*

1.1 Overview

For the three main topics in this thesis, the common underlying thread is that the standard techniques from each area do not apply directly; in each case, insights from probability are key to finding solutions to challenging problems from extremal combinatorics and structural graph theory. In general, the probabilistic method is a powerful technique for proving the existence of combinatorial objects with specified properties. The tremendous growth of combinatorics over the last century is due in part to its close relationship to probability theory, number theory, and theoretical computer science; results in each have applications and lead to breakthroughs in one another. Developing new techniques motivated by recent work in these areas produces some of the most beautiful results and most important contributions in combinatorics.

1.2 Viewing Extremal Problems through a Probabilistic Lens

The field of extremal combinatorics encompasses a wide variety of results. Fundamentally speaking, *extremal combinatorics* is the study of finite objects, such as graphs, sets, etc., that are *extremal*, meaning that they are maximal (or minimal) with respect to a certain property. In extremal graph theory, for instance, many classical results focus on finding how many edges a graph forbidding a fixed subgraph can have. One of the first extremal results is a theorem by Mantel [58] from 1907:

Theorem 1.2.1 (Mantel's Theorem)**.** *For any K_3-free graph G on n vertices, $e(G) \leq \frac{n^2}{4}$.*

Turán [67] generalized this result in 1941 as follows:

Theorem 1.2.2 (Turán's Theorem)**.** *For any K_{r+1}-free graph G on n vertices, $e(G) \leq (1 - \frac{1}{r})\frac{n^2}{2}$.*

Consider the complete multipartite graph $T(n,r)$ formed by partitioning n vertices into r parts, with sizes as equal as possible and connecting two vertices with an edge if and only if they belong to different parts. This graph, referred to as the *Turán graph*, contains no copy of K_{r+1} and has a total of $(1-\frac{1}{r})\frac{n^2}{2}$ edges, provided that r divides n. For Mantel's Theorem in particular this corresponds to a complete bipartite graph with parts of size $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$; this graph has a total of $\left\lfloor \frac{n^2}{4} \right\rfloor$ edges but no K_3. The *extremal function* $ex(n,H)$ is defined as the maximum number of edges in a graph on n vertices not containing a subgraph isomorphic to the graph H. In this notation, we see from above that $ex(n,K_{r+1}) = (1-\frac{1}{r})\frac{n^2}{2}$, provided that r divides n. The Erdős–Stone–Simonovits Theorem [31], [32] explores forbidding any fixed subgraph; note that for complete graphs, the chromatic number $\chi(K_{r+1}) = r+1$:

Theorem 1.2.3 (Erdős–Stone–Simonovits Theorem)**.** *For any graph H and any $\varepsilon > 0$, there exists n_0 so that for any $n \geq n_0$, $\left(1-\frac{1}{\chi(H)-1}-\varepsilon\right)\frac{n^2}{2} \leq \mathrm{ex}(n,H) \leq \left(1-\frac{1}{\chi(H)-1}+\varepsilon\right)\frac{n^2}{2}$.*

Once extremal results have been established, counting problems such as enumerating systems and describing their "typical structure" as the size of the underlying ground set tends to infinity are studied. Building on Mantel's Theorem, Erdős, Kleitman, and Rothschild [29] described typical K_3-free graphs:

Theorem 1.2.4. *Let $B(n)$ denote the set of bipartite graphs on n labeled vertices and $F(n)$ denote the set of all K_3-free graphs. Then $|F(n)| = ((1+o(1))\,|B(n)|$. Thus, almost all K_3-free graphs are bipartite.*

Erdős, Frankl, and Rödl [28] more generally were concerned with the asymptotic number of graphs without a fixed subgraph. More precisely, if $f_n(H)$ denotes the number of labeled H-free graphs on n labeled vertices, then Erdős, Frankl, and Rödl [28] showed that $f_n(H) = 2^{ex(n,H)+o\left(n^2\right)}$. More recently Balogh, Bollobás, and Simonovits [8] improved this to $f_n(H) = 2^{ex(n,H)+O\left(n^{2-c(H)}\right)}$ where $c(H)$ is a positive constant depending on H; furthermore Balogh, Bollobás, and Simonovits determined the typical structure of graphs without a fixed subgraph (and in fact determined the typical structure without fixed finite families of subgraphs).

A recent line of investigation is extending classical results to the so-called "sparse random setting". If $f_{n,m}(H)$ is the number of labeled H-free graphs on n labeled vertices with precisely m edges, then the following theorem, shown by Łuczak [53] and later by Balogh, Morris, and Samotij [10] using hypergraph containers, is a sparse version of the results above:

Theorem 1.2.5. *For every graph H and every positive δ, there exists a positive constant c such that the following holds. For every $n \in \mathbb{N}$, if $m \geq c \cdot n^{2-\frac{v(H)-2}{e(H)-1}}$, then*

$$\binom{ex(n,H)}{m} \leq f_{n,m}(H) \leq \binom{ex(n,H)+\delta n^2}{m}.$$

In extremal set theory, for $k \geq 2$ a k-uniform hypergraph $\mathcal{H}$ on vertex set $[n]$ is said to be *intersecting* if every pair of hyperedges shares at least one vertex; we say furthermore that a family is *trivial* if every edge in $\mathcal{H}$ contains some fixed vertex. The following result by Erdős, Ko, and Rado [30] shows that the trivial families are extremal:

Theorem 1.2.6 (Erdős–Ko–Rado Theorem)**.** *If $n \geq 2k$ and H is an intersecting k-uniform hypergraph on n vertices, then $e(H) \leq \binom{n-1}{k-1}$. Furthermore, for $n > 2k$ equality holds above only if H is trivial.*

As outlined here and appearing in Chapter 2, in joint work with Balogh, Das, Liu, and Sharifzadeh [9], we show that for intersecting families of k-uniform hypergraphs almost all intersecting families are trivial. We also enumerate such systems and explore the sparse random setting. As demonstrated in this thesis for permutations and vector spaces, this work fits into a more general framework that could be adapted for a variety of other settings provided the required extremal results are established.

Trivial Intersecting Families of Permutations

Let S_n denote the symmetric group on $[n]$. A family of permutations $\mathcal{F} \subseteq S_n$ is said to be *t-intersecting* if any two permutations in $\mathcal{F}$ agree on at least t indices; that is, for any $\sigma, \pi \in \mathcal{F}$,

$$|\sigma \cap \pi| = |\{i \in [n] : \sigma(i) = \pi(i)\}| \geq t.$$

When $t = 1$, we simply call such families *intersecting.* A natural example of a t-intersecting family $\mathcal{F} \subseteq S_n$ is a *trivial* t-intersecting family, where there is a fixed t-element subset $I \subseteq [n]$ and values $\{j_i : i \in I\}$ such that for every $\sigma \in \mathcal{F}$ and $i \in I$, $\sigma(i) = j_i$. Ellis, Friedgut, and Pilpel [27] proved that the extremal intersecting families are trivial:

Theorem 2.3.4. *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 trivial.*

In joint work with Balogh, Das, Liu, and Sharifzadeh [9], we determine the typical structure of t-intersecting families in S_n, showing that trivial families are not just extremal but also typical.

Theorem 2.3.6. *For any fixed $t \geq 1$ and n sufficiently large, almost all t-intersecting families of permutations in S_n are trivial, and there are $\left(\binom{n}{t}^2 t! + o(1)\right) 2^{(n-t)!}$ t-intersecting families.*

Additionally, we prove two results in the sparse random setting. First we obtain the following sparse extension of Theorem 2.3.4. Let $(S_n)_p$ denote the p-random subset of S_n, where each permutation in

S_n is included independently with probability p. Provided p is not too small, we show that with high probability the largest t-intersecting family in $(S_n)_p$ is trivial. Note that the work by Ellis, Friedgut, and Pilpel corresponds to the case when $p = 1$.

Theorem 2.3.8. *For fixed $t \geq 1$, n sufficiently large, and $p = p(n) \geq c \cdot \frac{n}{(n-t)!} \cdot 2^{2(n-t)} \cdot \log n$, with high probability every largest t-intersecting family in $(S_n)_p$ is trivial.*

In the second extension, we consider t-intersecting families of permutations of size m. Note that each trivial t-intersecting family contains $\binom{(n-t)!}{m}$ subfamilies of size m. The following result shows that, provided m is not too small, the number of non-trivial t-intersecting families of m permutations is a lower-order term.

Theorem 2.3.7. *For any fixed $t \geq 1$, n sufficiently large, and $n \cdot 2^{2n-2t+2} \cdot \log n \leq m \leq (n-t)!$, almost all t-intersecting families of m permutations in S_n are trivial.*

Trivial Intersecting Families of Hypergraphs

For $k \geq 2$ and $1 \leq t < k$, a k-uniform hypergraph $\mathcal{H}$ on vertex set $[n]$ is *t-intersecting* if every pair of edges shares at least t vertices. A family is *trivial* if every edge in $\mathcal{H}$ contains a fixed set of t vertices. The following is one of the most important results in the area of extremal combinatorics Erdős, Ko, and Rado [30]:

Theorem 2.4.1. *For $n \geq t + (k-t)\binom{k}{t}^3$, the largest t-intersecting k-uniform hypergraphs on $[n]$ have size at most $\binom{n-t}{k-t}$.*

The classic Erdős–Ko–Rado Theorem [30] and as well as the following result of Wilson [73] from 1984 (shown for $t > 14$ by Frankl [33] in 1978) demonstrate that the largest t-intersecting k-uniform hypergraphs are the trivial ones with $\binom{n-t}{k-t}$ edges.

Theorem 2.4.2. *For $n \geq (t+1)(k-t+1)$, the largest t-intersecting k-uniform hypergraphs on $[n]$ have size at most $\binom{n-t}{k-t}$.*

We show that just beyond this bound, the trivial t-intersecting k-uniform hypergraphs are typical.

Theorem 2.4.3. *Let n, $k = k(n) \geq 3$ and $t = t(n) \geq 1$ be integers such that $n \geq (t+1)(k-t+1) + \eta_{k,t}$, where*

$$\eta_{k,t} = \begin{cases} k + 8\ln k & \text{for } t = 1, \\ 12\ln k & \text{for } t = 2 \text{ and } k - t \geq 3, \\ 1 & \text{for } t \geq 3 \text{ and } k - t \geq 3 \\ 31 & \text{for } t \geq 2 \text{ and } k - t = 2, \\ 18k & \text{for } t \geq 2 \text{ and } k - t = 1. \end{cases}$$

Almost all t-intersecting k-uniform hypergraphs on $[n]$ are trivial, and there are $\left(\binom{n}{t} + o(1)\right) 2^{\binom{n-t}{k-t}}$ t-intersecting k-uniform hypergraphs.

For $\eta_{k,t} = 1$ (the value for which we have for most values of t and k) this is the best possible result. When $n = (t+1)(k-t+1)$, the largest non-trivial t-intersecting hypergraphs are actually as large as the trivial t-intersecting hypergraphs, and there are many more of them. For $n = (t+1)(k-t+1)$ almost every t-intersecting hypergraph is non-trivial.

The most natural, and arguably the most interesting, case to focus on is when $t = 1$. Theorem 2.4.3 gives the asymptotic number of intersecting hypergraphs when $n \geq 3k + 8\ln k$.

Theorem 2.1.4. *Let n and $k = k(n) \geq 3$ be integers such that $n \geq 3k+8\ln k$. Then there are $(n + o(1))\, 2^{\binom{n-1}{k-1}}$ intersecting k-uniform hypergraphs on $[n]$. Thus, almost all intersecting k-uniform hypergraphs are trivial.*

On the other hand, it is known that the trivial hypergraphs are the largest when $n \geq 2k$, and uniquely so when $n \geq 2k+1$. The following theorem, which we prove using spectral methods and the theory of graph containers, provides a slightly weaker result that covers the entire range.

Theorem 2.4. *For $k \geq 3$ and $n \geq 2k+1$, let $I(n,k)$ denote the number of intersecting k-uniform hypergraphs on $[n]$. Then*

$$\log I(n,k) = (1 + o(1)) \binom{n-1}{k-1}.$$

Let $\mathcal{H}^k(n,p)$ denote the p-random k-uniform hypergraph on $[n]$, in which every edge in $\binom{[n]}{k}$ is included independently at random with probability p. Balogh, Bohman, and Mubayi [7] initiated the study of intersecting hypergraphs in the sparse random setting. Among other results, they determined the size of the largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ when $k < n^{1/2-\varepsilon}$.

Similarly to the results for permutations, we are able to obtain a sparse version of the Erdős–Ko–Rado Theorem; this is a highly active area of study. Recently, Gauy, Hàn, and Oliveira [37] determined the asymptotic size of the largest intersecting family for all k and almost all p. Hamm and Kahn [46] obtained an exact result for $k < (\frac{1}{4} - c)(n \log n)^{1/2}$ for some small constant c and $p \gg \binom{n-k}{k}^{-1}$, showing that with high probability every largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ is trivial. We prove that the same holds for k as large as linear in n, provided p is somewhat larger.

Theorem 2.4.4. *For $3 \leq k \leq \frac{n}{4}$, if*

$$p = p(n) \geq c \cdot n \cdot \frac{\binom{2k}{k}\binom{n}{k}}{\binom{n-k}{k}^2} \cdot \log\left(\frac{ne}{k}\right),$$

then with high probability every largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ is trivial.

Hamm and Kahn [40] also studied the case $n = 2k + 1$ and $p = 1 - c$ for some constant $c > 0$.

Trivial Intersecting Families of Vector Spaces

Let V be an n-dimensional vector space over a finite field $\mathbb{F}_q$. In this context, a family $\mathcal{F}$ of k-dimensional subspaces of V is *intersecting* if $\dim(F_1 \cap F_2) \geq 1$ for all pairs of subspaces $F_1, F_2 \in \mathcal{F}$. The number of k-dimensional subspaces in V is given by the Gaussian binomial coefficient

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

Hsieh [42] proved an Erdős–Ko–Rado-type Theorem for vector spaces:

Theorem 2.5.1. *For $n \geq 2k + 1$, any intersecting family $\mathcal{F}$ of k-dimensional subspaces of V has size at most $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$; and equality holds only if $\mathcal{F}$ is trivial.*

Furthermore, the only constructions achieving the maximum size consisting of all k-dimensional subspaces through a given 1-dimensional subspace, the natural notion of *trivial* in this context.

Here we determine that the typical structure of intersecting families of subspaces is trivial as well.

Theorem 2.5.2. *If $k \geq 2$, and either $q = 2$ and $n \geq 2k + 2$ or $q \geq 3$ and $n \geq 2k + 1$, almost all intersecting families of k-dimensional subspaces of $\mathbb{F}_q^n$ are trivial, and there are $\left(\begin{bmatrix} n \\ 1 \end{bmatrix}_q + o(1) \right) 2^{\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q}$ intersecting families.*

1.3 Viewing Structural Problems through a Probabilistic Lens

Structural graph theory is the study of when graphs have certain structural properties. As observed above, moving extremal problems to a random environment can be a natural further direction; the same can be said about certain structural problems.

Star Decompositions of Random Regular Graphs

As Barát and Thomassen [11] note, decompositions of the edges of a graph G into copies of a small fixed subgraph can be related to orientations with certain requirements. For instance, given a 4-regular planar graph G, an orientation in G with out-degrees 0 or 3 corresponds to an S_3-*decomposition* of G (a decomposition of the edges of G into copies of S_3). A graph is d-edge-connected if one must remove at least d edges in order to disconnect the graph. Barát and Thomassen [11] asked if every 4-edge-connected, 4-regular graph has an orientation with out-degrees 0 or 3. Barát and Thomassen observed that the answer is no and posed the following conjecture:

Conjecture 3.1.7. *Every 4-edge-connected, 4-regular planar graph has an orientation with out-degrees 0 or 3.*

Interestingly, a typical d-regular graph is d-edge-connected [14]. We translate this structural problem to the setting of random d-regular graphs:

Theorem 3.1.8. *A random 4-regular graph on n vertices has an orientation with out-degrees 0 or 3 asymptotically almost surely, provided that $2n$ is divisible by 3.*

Although this appears to be a straightforward application of the second moment method [4], standard probabilistic techniques do not work here; if $Y = Y(n)$ is the number of orientations of a random 4-regular graph on n vertices with out-degrees 0 or 3, then $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \sim \sqrt{\frac{3}{2}} > 0$. Instead we use the Small Subgraph Conditioning Method [61] to show $Y > 0$ asymptotically almost surely. When this powerful method works, by conditioning on the small subgraph counts, we are able to alter $\mathbb{E}[Y]$ by a constant factor and conclude that $Y > 0$ asymptotically almost surely.

Highly Connected, Spanning, Bipartite Subgraphs

A graph is said to be *k-edge-connected* if one must remove at least k edges in order to disconnect the graph. A related concept is graph G is said to be *k vertex-connected* if one must remove at least k vertices from $V(G)$ in order to disconnect the graph (or to remain with one single vertex). We also let $\kappa(G)$ denote the minimum integer k for which G is k-connected.

In 1988 Thomassen [66] observed that highly edge-connected graph contain large a highly edge-connected bipartite subgraph:

Theorem 4.1.2. *If G is a graph which is at least $(2k-1)$-edge-connected, then G contains a spanning, bipartite subgraph $H \subseteq G$ such that k-edge-connected.*

He conjectured that the same should hold for vertex-connectivity; the following appears as Conjecture 7 in Thomassen's [65] survey paper from 1989:

Conjecture 4.1.3. *For all k, there exists a function $f(k)$ such that for all graphs G, if the vertex-connectivity $\kappa(G) \geq f(k)$, then there exists a spanning, bipartite subgraph $H \subseteq G$ such that $\kappa(H) \geq k$.*

We show that this conjecture is true up to a $\log n$ factor by carefully constructing an auxiliary digraph.

Theorem 4.3.1. *For all k and n, and for every graph G on n vertices the following holds. If $\kappa(G) > 10^{10}k^3 \log n$, then there exists a spanning, bipartite subgraph $H \subseteq G$ such that $\kappa(H) \geq k$.*

1.4 Basic Definitions

A *graph* G is a pair $(V(G), E(G))$ consisting of a set $V(G)$ of *vertices* along with a set $E(G)$ of *edges* which consists of 2-element subsets of $V(G)$; the pair of vertices in each edge are unordered. The *order* of a graph G is the cardinality of the vertex set $|V(G)|$ denoted here as $v(G)$. Similarly the *size* of a graph G is the cardinality of the edge set $|E(G)|$ denoted here as $e(G)$. Two vertices $u, v \in V(G)$ are said to be *adjacent*, denoted $u \sim v$, uv, or vu, if $\{u, v\} \in E(G)$. An edge and a vertex on that edge are said to be *incident*. The type of graph defined above is referred to in the literature as a *simple graph* because there are no loops (a *loop* is an edge $u \sim u$) or multiple edges (several edges $u \sim v$). A graph allowing loops or multiple edges is referred to as a *multigraph*.

The *adjacency matrix* of G, say A, is an $v(G) \times v(G)$ matrix where element A_{ij} is the number of edges between vertex v_i and vertex v_j. Note that the adjacency matrix of G is an $v(G) \times v(G)$ symmetric matrix with all real entries (in fact this matrix is Hermitian) and therefore this matrix has $v(G)$ real eigenvalues. The eigenvalues of the adjacency matrix of G are often referred to as the *eigenvalues of the graph G*. An *isomorphism* from a graph G to a graph H is a bijection $f : V(G) \to V(H)$ preserving the adjacency relation. Note that $uv \in E(G)$ if and only if $f(u)f(v) \in E(H)$. Two graphs G_1 and G_2 with adjacency matrices A_1 and A_2 respectively are said to be *isomorphic* if and only if there exists a permutation matrix P such that $PA_1P^{-1} = A_2$.

The *degree* of a vertex is the number of edges incident to that vertex (for multigraphs we count loops twice). An *isolated vertex* has degree 0. A graph is said to *d-regular* if all vertices have the degree d. The *average degree* of a graph is the sum of all the degrees divided by the number of vertices. A *path* is a graph whose vertices may be linearly ordered so that two vertices are adjacent if and only if they appear consecutively in the ordering. A graph is said to be *connected* is there is a path between any pair of vertices. A graph H is said to be a *subgraph* of a graph G, denoted $H \subseteq G$, if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. A maximal connected subgraph is referred to as a *component* of the graph. A graph G is said to be *k-vertex-connected* if one must remove at least k vertices in order to disconnect the graph (or to remain with one single vertex). A graph is said to be *k-edge-connected* if one must remove at least k edges in order to disconnect the graph. A *vertex cover* is a subset $S \subseteq V(G)$ such that every edge in G is incident to at least one vertex in S. The minimum number of vertices whose deletion disconnects a graph G is the *vertex-connectivity*, denoted $\kappa(G)$. A *separating set* is a subset of the vertex set whose deletion makes the graph disconnected.

An a *independent set* I, sometimes known as a *stable set*, in a graph G is a subset of $V(G)$ that forms no edges. The *independence number* of a graph G, denoted $\alpha(G)$, is number of vertices in a largest independent set of G. Likewise, a *matching* M in a graph G is a subset of $E(G)$ that share no vertices. A matching is said to be *perfect* if every vertex is incident with an edge of M.

A *coloring* of a graph G is an assignment of labels to the vertices of G such no two adjacent vertices receive the same label. By definition, every color class in a valid coloring must be an independent set. A *k-coloring* of a graph G is a coloring using k-colors. The *chromatic number* of a graph G, denoted $\chi(G)$, is the minimum value k such that G has a valid k-coloring.

The *complete graph* on n vertices, denoted K_n, is the graph where every pair of distinct vertices is connected by exactly one edge. A *cycle* on n edges, denoted C_n, has n edges and there is a cyclic order of the vertices so that two vertices are adjacent if and only if they appear consecutively in this ordering. *Bipartite graphs* are the graphs with chromatic number at most 2. A graph is bipartite if and only if it does not contain an odd cycle. The vertices of a bipartite graph can be partitioned into two disjoint sets such that no two vertices within the same set are adjacent. A *complete bipartite graph* is a bipartite graph such that every pair of graph vertices in the two parts of the partition are adjacent; this graph is denoted $K_{\ell,k}$ where ℓ and k are the number of vertices in the two disjoint parts. A bipartite connected graph with no cycles is said to be a *tree*. In a tree, the vertices of degree 1 are called *leaves*. The graph $S_k \simeq K_{1,k}$ is an example of a tree and is referred to as the *star* with k leaves in the literature. A subgraph $F \subseteq G$ is said to be *spanning* if $V(F) = V(G)$. For a fixed subgraph $F \subseteq G$, an *F-decomposition* is a partition of $E(G)$ into disjoint copies of F. The degree of a vertex $v \in V(G)$ in a subgraph $H \subseteq G$ is denoted $d_H(v)$.

A *hypergraph* $\mathcal{H}$ is a pair $(V(\mathcal{H}), E(\mathcal{H}))$ consisting of a set $V(\mathcal{H})$ of *vertices* along with a set $E(\mathcal{H})$ of *hyperedges* which consists of subsets of $V(\mathcal{H})$. Similarly, the *order* of a hypergraph $\mathcal{H}$ is the cardinality of the vertex set $|V(\mathcal{H})|$ denoted here as $v(\mathcal{H})$, and the *size* of a hypergraph $\mathcal{H}$ is the cardinality of the edge set $|E(\mathcal{H})|$ denoted here as $e(\mathcal{H})$. The vertex set is sometimes referred to as the *ground set* and $[n] := \{1, 2, \ldots n\}$ will be the ground set of many hypergraphs in this thesis. If all of the hyperedges have exactly k elements from $V(\mathcal{H})$, then $\mathcal{H}$ is referred to as k-uniform. Here the notion of a graph above corresponds with 2-uniform hypergraphs. The k-uniform hypergraph on n having all possible k-element edges is denoted $\binom{[n]}{k}$; this hypergraph has size $\binom{n}{k} := \frac{n!}{k!(n-k)!}$, where the factorial $n! := 1 \cdot 2 \cdot \ldots \cdot n$. On the other hand, we define the *falling factorial* $[n]_k := n \cdot (n-1) \cdot \ldots \cdot (n-(k-1))$. We let $\mathcal{H}^k(n,p)$ denote the p-random k-uniform hypergraph on $[n]$ in which every edge in $\binom{[n]}{k}$ is included independently with probability p.

As before, an *independent set* I in a hypergraph $\mathcal{H}$ is a subset of $V(\mathcal{H})$ that forms no edges. The *independence number* of a hypergraph $\mathcal{H}$, denoted $\alpha(\mathcal{H})$, is number of vertices in a largest independent set of $\mathcal{H}$. Likewise, the *degree* of a vertex in a hypergraph is the number of hyperedges incident to that vertex, and a hypergraph is said to *d-regular* if all vertices have the degree d.

An *orientation* of a graph is an assignment of exactly one direction to each of the edges. A *directed graph* or *digraph*, on the other hand, allows bidirectional edges in addition. For a directed graph D and a vertex $v \in V(D)$ we let $d_D^+(v)$ denote the out-degree of v. We let $U(D)$ denote the *underlying graph* of D, that is the graph obtained by ignoring the directions in D and merging any multiple edges. For both digraphs and orientations, for a vertex v, $\delta^+(v)$ denotes the set of edges out of v whereas $\delta^-(v)$ denotes the set of edges into v.

A *planar graph* is a graph that can be embedded in the plane without crossings. A particular planar embedding is referred to as a *plane graph*. The *dual graph* of a plane graph G is the graph with a vertex for every face of G and an edge between two vertices when the corresponding faces are separated from each other by an edge.

For a set S, the *power set* of S, denoted $\mathcal{P}(S)$, is the set of all subsets of S, including the empty set $\emptyset$ and S itself. Let $\varphi(n)$ be a positive function. Little-o notation, denoted $o(x)$, is a Landau symbol describing the asymptotic behavior of a function, and $o(f) = \{g : |g(x)|/|f(x)| \to 0\}$. Whereas with big-$O$ notation, $O(f) = \{g : \exists c, a \in \mathbb{R} \text{ such that } |g(x)| \leq c \cdot |f(x)| \text{ for } x > a\}$. Big-$\Omega$ notation is the inverse of big-O notation:

$$f(n) \in O(g(n)) \text{ if and only if } g(n) \in \Omega(f(n)).$$

Throughout this thesis,

$$e := \sum_{n=0}^{\infty} \frac{1}{n!}$$

is the constant referred to as Euler's number, and ln denotes the *natural logarithm*, the logarithm to the base e. Unless otherwise stated, here log denotes the logarithm to the base 2.

For standard graph theory definitions and background see Diestel [23] or West [70], [71], and [72]. For an exposition on random graphs see Bollobás [14]. The classical text for introducing the probabilistic method is Alon and Spencer [4].

Chapter 2

The Typical Structure of Intersecting Families

2.1 Introduction

A fundamental question in extremal combinatorics asks: how large can a system be under certain restrictions? Once resolved, this can be strengthened by enumerating such systems and describing their "typical structure" as the size of the underlying ground set tends to infinity. In extremal graph theory, for instance, this study was initiated by Erdős, Kleitman, and Rothschild [29] in 1976. Their results, explained in further detail below, have inspired a great deal of research over the years by motivating the extension of classical theorems from many areas in this manner; e.g., see [8] and [28]. A core topic explored in this thesis is determining the typical structure of intersecting families of discrete systems in various settings[1].

2.1.1 Triangle-Free Graphs

One of the most fundamental results in extremal graph theory is this theorem by Mantel [58] from 1907:

Theorem 2.1.1 (Mantel's Theorem)**.** *If a graph G on n vertices contains no triangle (cycle on three vertices), then it has at most $\frac{n^2}{4}$ edges.*

To see that this is best possible, consider partitioning n isolated vertices into two disjoint sets, one of size $\left\lfloor \frac{n}{2} \right\rfloor$ and the other of size $\left\lceil \frac{n}{2} \right\rceil$, and adding all possible edges from one set to the other. This complete bipartite graph has a total of $\left\lfloor \frac{n^2}{4} \right\rfloor$ edges but no triangle.

Erdős, Kleitman, and Rothschild [29] describe what a typical triangle-free graph looks like:

Theorem 2.1.2. *Let $B(n)$ denote the set of bipartite graphs on n labeled vertices and $F(n)$ denote the set of all triangle-free graphs. Then $|F(n)| = ((1 + o(1)) |B(n)|$. Thus, almost all triangle-free graphs are bipartite.*

Therefore, counting the number of bipartite graphs (an easy problem) actually helps to count the total number of triangle-free graphs (a harder problem) because the cardinality of these two classes of graphs is asymptotically equal.

[1]Some of this work appeared in the Journal of Combinatorial Theory, Series A in 2015; see [9].

2.1.2 Trivial Intersecting Families

The study of intersecting families of various discrete structures is a central and very active area of research in extremal set theory. A k-uniform hypergraph on n vertices is said to be *intersecting* if any two hyperedges share at least one vertex and is said furthermore to be *trivial* if all hyperedges share at least one vertex. A natural question in this setting is: how many hyperedges can an intersecting k-uniform hypergraph on n vertices have? A large intersecting k-uniform hypergraph on $[n]$ can be constructed by designating one vertex as special and forming hyperedges by selecting all possible k-element subsets containing that vertex. After fixing one vertex, there are $n-1$ other vertices from which to pick the remaining $k-1$ vertices for each hyperedge; thus, these maximal trivial intersecting k-uniform hypergraphs on $[n]$ have $\binom{n-1}{k-1}$ hyperedges in total.

One of the oldest and most influential results in extremal combinatorics appears in the seminal 1961 paper of Erdős, Ko, and Rado [30] and shows for $n > 2k$ that the construction above is extremal:

Theorem 2.1.3 (Erdős–Ko–Rado Theorem). *If $n \geq 2k$ and H is an intersecting k-uniform hypergraph on n vertices, then $e(H) \leq \binom{n-1}{k-1}$. Furthermore, for $n > 2k$ equality holds above only if H is trivial.*

In the case when $n = 2k$ another intersecting k-uniform hypergraph with $\frac{1}{2}\binom{2k}{k} = \binom{2k-1}{k-1} = \binom{n-1}{k-1}$ hyperedges can be constructed by taking exactly one hyperedge from each complementary pair E and E^c in $\binom{[2k]}{k}$. When $n < 2k$, any two k-element subsets must share at least one element so the maximum number of hyperedges in an intersecting k-uniform hypergraph on n vertices is $\binom{n}{k}$.

In joint work with Balogh, Das, Liu, and Sharifzadeh [9], we show that the trivial intersecting k-uniform hypergraphs are typical:

Theorem 2.1.4. *Let n and $k = k(n) \geq 3$ be integers such that $n \geq 3k+8\ln k$. Then there are $(n+o(1))\, 2^{\binom{n-1}{k-1}}$ intersecting k-uniform hypergraphs on $[n]$. Thus, almost all intersecting k-uniform hypergraphs are trivial.*

A great deal of modern research is still devoted to proving analogous versions of the Erdős–Ko–Rado Theorem in other settings, e.g. [22], [27], [39], [42], [73]. Our proof that the typical structure of intersecting uniform hypergraphs is trivial provides a general framework that works in a variety of other settings. The remainder of this chapter is organized as follows. First is an outline of the proof method which holds for a number of similar problems, followed by the proofs of related results for permutations, uniform hypergraphs, and vector spaces, including both the versions for t-intersecting families as well as an exploration of the sparse random setting. The chapter concludes with an exposition of an alternate approach using graph containers followed by some open questions.

2.2 General Framework

The general framework for analyzing the typical structure of intersecting families of discrete systems in various settings consists of two stages. An intersecting family is said to be *maximal* if it is intersecting and is not contained in a larger intersecting family. First we obtain a strong upper bound on the number of maximal intersecting families. Next we combine this upper bound with known extremal and stability results in order to bound the number of non-trivial intersecting families; in each setting, this allows us to conclude that the trivial intersecting families are typical.

This section outlines the main ideas behind these steps, using intersecting k-uniform hypergraphs on $[n]$ as a running example. Section 2.3 will explore the setting of intersecting families of permutations, Section 2.4 will return to intersecting k-uniform hypergraphs, and Section 2.5 will discuss analogous results for intersecting families of vector spaces. In subsequent sections, we will also generalize these results by proving analogous results for t-intersecting families and study what happens when we shift into the sparse random setting.

For the sake of clarity, many calculations will be omitted from this section and carried out in greater detail in Section 2.4.

2.2.1 Maximal Intersecting Families

Given a family of sets $\mathcal{F} \subseteq \binom{[n]}{k}$, we introduce the notion of the family of all sets intersecting every set in $\mathcal{F}$,

$$\mathcal{I}(\mathcal{F}) := \left\{ G \in \binom{[n]}{k} : \forall F \in \mathcal{F},\ G \cap F \neq \emptyset \right\}.$$

This family has some useful properties. We note that $\mathcal{F}$ is an intersecting family if and only if $\mathcal{F} \subset \mathcal{I}(\mathcal{F})$, and furthermore, $\mathcal{F}$ is a maximal intersecting family if and only if $\mathcal{F} = \mathcal{I}(\mathcal{F})$. Given a maximal intersecting family $\mathcal{F}$, we refer to $\mathcal{G} \subseteq \mathcal{F}$ as a *generating set* if $\mathcal{F} = \mathcal{I}(\mathcal{G})$.

Let $\mathcal{F}$ be a maximal intersecting family and $\mathcal{F}_0 = \{F_1, F_2, \ldots, F_s\} \subset \mathcal{F}$ be a minimal generating set of $\mathcal{F}$. This is well defined as $\mathcal{F} = \mathcal{I}(\mathcal{F})$ and therefore $\mathcal{F}$ generates itself. Observe that, by the minimality of $\mathcal{F}_0$, we have $\mathcal{F} = \mathcal{I}(\mathcal{F}_0) \subsetneq \mathcal{I}(\mathcal{F}_0 \setminus \{F_i\})$ for each $1 \leq i \leq s$ because $\mathcal{I}(\mathcal{F}_0) \subseteq \mathcal{I}(\mathcal{F}_0 \setminus \{F_i\})$ but $\mathcal{F} \neq \mathcal{I}(\mathcal{F}_0 \setminus \{F_i\})$. Hence for each i we can find some set $G_i \in \mathcal{I}(\mathcal{F}_0 \setminus \{F_i\}) \setminus \mathcal{F}$. Because $G_i \in \mathcal{I}(\mathcal{F}_0 \setminus \{F_i\})$, we have $G_i \cap F_j \neq \emptyset$ for all $F_j \in \mathcal{F}_0 \setminus \{F_i\}$, and therefore $G_i \cap F_j \neq \emptyset$ for all $i \neq j$; on the other hand, because $G_i \notin \mathcal{F} = \mathcal{I}(\mathcal{F}_0)$, we must have that $G_i \cap F_i = \emptyset$ for all i. To bound the size of $\mathcal{F}_0$, we may now apply Frankl's skew version [34] of the celebrated Bollobás Set-Pairs Inequality [15]:

Theorem 2.2.1 (Skew-Symmetric Bollobás Set-Pairs Inequality). *Let $A_1, \dots, A_m$ be sets of size a and $B_1, \dots, B_m$ be sets of size b such that $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ for every $1 \leq i < j \leq m$. Then $m \leq \binom{a+b}{a}$.*

Given the collections of k-element subsets $\{F_1, F_2, \dots, F_s\}$ and $\{G_1, G_2, \dots, G_s\}$ from above, we construct a system of set-pairs $\{(A_i, B_i)\}_{i=1}^{2s}$ as follows. For $1 \leq i \leq s$, let $A_i = F_i$ and $B_i = G_i$, and for $s+1 \leq i \leq 2s$, let $A_i = G_{i-s}$ and $B_i = F_{i-s}$. From the argument above $G_i \cap F_i = \emptyset$, and thus, $A_i \cap B_i = \emptyset$ for every $1 \leq i \leq 2s$. Furthermore, for all $1 \leq i, j \leq s$ with $i \neq j$, we know that $G_i \cap F_j \neq \emptyset$. Therefore, $A_i \cap B_j \neq \emptyset$ for all $1 \leq i, j \leq s$ with $i \neq j$, and likewise $A_i \cap B_j \neq \emptyset$ for all $s+1 \leq i, j \leq 2s$ with $i \neq j$. Because $\mathcal{F}$ is intersecting then so is $\mathcal{F}_0 \subseteq \mathcal{F}$ and $A_i \cap B_j \neq \emptyset$ for all $1 \leq i \leq s$ and $s+1 \leq j \leq 2s$ because $F_i \cap F_{j-s} \neq \emptyset$. Thus, $A_i \cap B_j \neq \emptyset$ for every $1 \leq i < j \leq 2s$. Thus, the set pairs $\{(A_i, B_i)\}$ satisfy the conditions of Theorem 2.2.1, and we may deduce that $2s \leq \binom{2k}{k}$. Thus, $|\mathcal{F}_0| = s \leq \frac{1}{2}\binom{2k}{k}$.

The fact that every maximal intersecting family admits a small generating set allows us to bound the number of maximal intersecting families.

Proposition 2.2.2. *The number of maximal intersecting k-uniform hypergraphs on $[n]$ is at most*

$$\sum_{i=0}^{\frac{1}{2}\binom{2k}{k}} \binom{\binom{n}{k}}{i} \leq \binom{n}{k}^{\frac{1}{2}\binom{2k}{k}}.$$

Proof. We map each maximal intersecting k-uniform hypergraph $\mathcal{F}$ to a minimal generating set $\mathcal{F}_0 \subset \mathcal{F}$. Although $\mathcal{F}_0$ may not be unique, because $\mathcal{F} = \mathcal{I}(\mathcal{F}_0)$, this map from $\mathcal{F}$ to $\mathcal{F}_0$ is injective. Using the Skew-Symmetric Bollobás Set-Pairs Inequality as above, we see that $|\mathcal{F}_0| \leq \frac{1}{2}\binom{2k}{k}$, and hence, a bound for the number of maximal intersecting k-uniform hypergraphs is provided by the number of sets of at most $\frac{1}{2}\binom{2k}{k}$ edges. This is precisely the summation above. $\square$

2.2.2 Non-trivial Intersecting Families

We now would like to combine our upper bound on the number of maximal families with other known results in order to bound the total number of non-trivial intersecting families. We can then conclude that the trivial intersecting families are typical. The crux of the argument relies on two simple observations. The first point to observe is that any subset of a trivial intersecting family is also in fact itself a trivial intersecting family. The second observation is that every non-trivial intersecting family must be a subset of some maximal non-trivial family.

As follows, obtaining results that show both that the trivial intersecting families are the largest intersecting families and bounds the size of the largest non-trivial family away from the size of the largest trivial intersecting families is crucial. Fortunately this has been studied for a variety of discrete structures. The following lemma gives sufficient conditions for the trivial families to be typical and will be applicable in all of our settings:

Lemma 2.2.3. *Let N_0 denote the size of the largest trivial intersecting family, and let N_1 denote the size of the largest non-trivial intersecting family. Suppose further that there are at most M maximal intersecting families. Provided*

$$\log M + N_1 - N_0 \to -\infty, \tag{2.1}$$

almost all intersecting families are trivial. Moreover, if m is such that

$$\log M - m \log \left(\frac{N_0}{N_1} \right) \to -\infty, \tag{2.2}$$

then almost all intersecting families of size m are trivial.

Proof. Because a largest trivial intersecting family has size N_0, and all of the subfamilies of the family are also trivial intersecting families, there must be at least 2^{N_0} trivial families. On the other hand, every non-trivial intersecting family is a subset of a maximal non-trivial intersecting family. Each maximal non-trivial intersecting family has size at most N_1, and thus at most 2^{N_1} subfamilies. Because there are at most M maximal intersecting families, the number of non-trivial intersecting families is at most $M2^{N_1}$. The proportion of non-trivial families is thus at most $M2^{N_1}/2^{N_0}$, which tends to 0 by (2.1). Hence, given (2.1), almost all intersecting families are trivial.

For the second claim, observe that the number of trivial intersecting subfamilies of size m is at least $\binom{N_0}{m}$ by considering all possible subfamilies of one fixed trivial intersecting family. On the other hand, each non-trivial intersecting family has at most $\binom{N_1}{m}$ subfamilies of size m, and hence there are at most $M\binom{N_1}{m}$ non-trivial intersecting families of size m. We can thus bound the proportion of intersecting families of size m that are non-trivial by

$$M \binom{N_1}{m} / \binom{N_0}{m} \leq M \left(\frac{N_1}{N_0} \right)^m,$$

which tends to 0 by (2.2); thus, almost all intersecting families of size exactly m are trivial as well. □

As mentioned before in Theorem 2.1.3, the Erdős–Ko–Rado Theorem [30] states that for $n > 2k$, the largest intersecting k-uniform hypergraphs over $[n]$ are trivial, having size $\binom{n-1}{k-1}$. A natural next question is: if we throw out the trivial families, then how large can an intersecting family be? A stability result was given for intersecting k-uniform hypergraphs over $[n]$ by Hilton and Milner [41] in 1967:

Theorem 2.2.4 (Hilton–Milner Theorem). *If $n > 2k$ and H is a non-trivial intersecting k-uniform hypergraph on n vertices, then $e(H) \le \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$.*

These two results provide the values of N_0 and N_1 respectively, whereas M is given by Proposition 2.2.2.

Finally, having determined that almost all intersecting families are trivial, we still must count the number of such families. The following lemma shows when the union bound over all maximal trivial families gives an asymptotical correct result.

Lemma 2.2.5. *Let T denote the number of maximal trivial intersecting families, and suppose they all have the same size N_0. Suppose further that two distinct maximal families can have at most N_2 members in common. Provided*

$$2 \log T + N_2 - N_0 \to -\infty, \tag{2.3}$$

the number of trivial intersecting families is $(T + o(1))\, 2^{N_0}$.

Proof. Suppose $\mathcal{F}_1, \ldots, \mathcal{F}_T$ are the maximal trivial intersecting families. Every trivial family is a subset of some $\mathcal{F}_i$, and hence the collection of trivial families is given by $\bigcup_{i=1}^{T} \mathcal{P}(\mathcal{F}_i)$. The Bonferroni Inequalities state that, for any sets $\mathcal{G}_1, \ldots, \mathcal{G}_m$,

$$\sum_{i=1}^{m} |\mathcal{G}_i| - \sum_{i<j} |\mathcal{G}_i \cap \mathcal{G}_j| \le \left| \bigcup_{i=1}^{m} \mathcal{G}_i \right| \le \sum_{i=1}^{m} |\mathcal{G}_i| \, .$$

Applying these inequalities with $\mathcal{G}_i = \mathcal{P}(\mathcal{F}_i)$ for $1 \le i \le m = T$, we have $|\mathcal{G}_i| = |\mathcal{P}(\mathcal{F}_i)| = 2^{N_0}$ and $|\mathcal{G}_i \cap \mathcal{G}_j| = |\mathcal{P}(\mathcal{F}_i \cap \mathcal{F}_j)| \le 2^{N_2}$. This gives

$$\sum_{i} |\mathcal{G}_i| = T \cdot 2^{N_0} \quad \text{and} \quad \sum_{i<j} |\mathcal{G}_i \cap \mathcal{G}_j| \le 2^{N_2} \binom{T}{2} < 2^{2 \log T + N_2 - N_0} \cdot 2^{N_0} = o\left(2^{N_0}\right),$$

from which the result follows. $\square$

This general framework, combined with the appropriate extremal and stability theorems, allows us to obtain our results, although minor modifications are required in the various settings. In many settings, these required extremal and stability results are well studied and what remains is to obtain a strong upper bound on the number of maximal intersecting families. In the following sections in this chapter we present calculations and describe the necessary changes needed to apply Lemmas 2.2.3 and 2.2.5 for permutations, uniform hypergraphs, and vector spaces.

2.3 Permutations

Let S_n denote the symmetric group on $[n]$. A family of permutations $\mathcal{F} \subseteq S_n$ is said to be *intersecting* if any two permutations in $\mathcal{F}$ agree on some index; that is, for any $\sigma, \pi \in \mathcal{F}$,

$$|\sigma \cap \pi| = |\{i \in [n] : \sigma(i) = \pi(i)\}| \geq 1.$$

A example of a family of permutations $\mathcal{F} \subseteq S_n$ that is intersecting can be obtained by fixing $i \in [n]$ and (some not necessarily distinct value) $j \in [n]$ such that for every $\sigma \in \mathcal{F}$, $\sigma(i) = j$; this a *trivial* intersecting family in the setting of permutations.

Deza and Frankl [22] proved an Erdős–Ko–Rado–type result (recall Theorem 2.1.3) for intersecting families of permutations in 1977:

Theorem 2.3.1. *For n sufficiently large, an intersecting family $\mathcal{F} \subseteq S_n$ has size at most $(n-1)!$.*

In 2003 Cameron and Ku [17] and independently Larose and Malvenuto [52] showed that here equality holds only if $\mathcal{F}$ is trivial. Ellis [25] showed a Hilton–Milner–type result (recall Theorem 2.2.4) for intersecting families of permutations:

Theorem 2.3.2. *For n sufficiently large, the largest non-trivial intersecting family $\mathcal{F} \subseteq S_n$ has size at most $\left(1 - \frac{1}{e} + o(1)\right)(n-1)!$.*

We determine the typical structure of intersecting families in S_n using the general framework outlined in Section 2.2, showing that trivial intersecting families are not just extremal but also typical:

Theorem 2.3.3. *For n sufficiently large, almost all intersecting families of permutations in S_n are trivial, and there are $\left(n^2 + o(1)\right) 2^{(n-1)!}$ intersecting families.*

A family of permutations $\mathcal{F} \subseteq S_n$ is said to be *t-intersecting* if any two permutations in $\mathcal{F}$ agree on at least t indices; that is, for any $\sigma, \pi \in \mathcal{F}$,

$$|\sigma \cap \pi| = |\{i \in [n] : \sigma(i) = \pi(i)\}| \geq t.$$

This is a natural generalization of intersecting; consider $t = 1$. The notion of trivial intersecting families can also be generalized; a *trivial* t-intersecting family has a fixed t-element subset $I \subseteq [n]$ and values $\{j_i : i \in I\}$ such that for every $\sigma \in \mathcal{F}$ and $i \in I$, $\sigma(i) = j_i$. In 2011 Ellis, Friedgut, and Pilpel [27] proved an Erdős–Ko–Rado–type result for t-intersecting families of permutations:

Theorem 2.3.4. *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 trivial.*

Ellis [25] showed an analogous Hilton–Milner–type result:

Theorem 2.3.5. *For n sufficiently large with respect to t, a largest non-trivial t-intersecting family $\mathcal{F} \subseteq S_n$ has size $(1 - 1/e + o(1))\,(n-t)!$.*

Our general framework extends to t-intersecting families in S_n as well, showing that trivial t-intersecting families are typical in this setting.

Theorem 2.3.6. *For any fixed $t \geq 1$ and n sufficiently large, almost all t-intersecting families of permutations in S_n are trivial, and there are $\left(\binom{n}{t}^2 t! + o(1)\right) 2^{(n-t)!}$ t-intersecting families.*

Theorem 2.3.3 follows as a corollary.

Additionally, we prove two extensions of Theorem 2.3.6 in the sparse setting. In the first we consider t-intersecting families of permutations of size m. Note that each trivial t-intersecting family contains $\binom{(n-t)!}{m}$ subfamilies of size m. The following result shows that, provided m is not too small, the number of non-trivial t-intersecting families of m permutations is a lower-order term.

Theorem 2.3.7. *For any fixed $t \geq 1$, n sufficiently large, and $n2^{2n-2t+2} \log n \leq m \leq (n-t)!$, almost all t-intersecting families of m permutations in S_n are trivial.*

Secondly we obtain the following sparse extension of the result of Ellis, Friedgut, and Pilpel [27]. Let $(S_n)_p$ denote the p-random subset of S_n, where each permutation in S_n is included independently with probability p. Provided p is not too small, we show that with high probability the largest t-intersecting family in $(S_n)_p$ is trivial. Note that Theorem 2.3.4 of Ellis, Friedgut, and Pilpel corresponds to the case when $p = 1$.

Theorem 2.3.8. *For fixed $t \geq 1$, n sufficiently large, and $p = p(n) \geq \frac{800n2^{2n-2t} \log n}{(n-t)!}$, with high probability every largest t-intersecting family in $(S_n)_p$ is trivial.*

Following the framework introduced in Section 2.2, we first bound the number of maximal t-intersecting families of permutations, and then deduce from this Theorems 2.3.6, 2.3.7, and 2.3.8.

Proposition 2.3.9. *For any $n \geq t \geq 1$, the number of maximal t-intersecting families in S_n is at most*

$$\sum_{i=0}^{\frac{1}{2}\binom{2n-2t+2}{n-t+1}} \binom{n!}{i} < n^{n2^{2n-2t+1}}.$$

Proof. Following the proof of Proposition 2.2.2, for a maximal t-intersecting family $\mathcal{F} \subset S_n$, we define

$$\mathcal{I}(\mathcal{F}) = \{\pi \in S_n : \forall \sigma \in \mathcal{F},\ |\pi \cap \sigma| \geq t\}.$$

Let $\mathcal{F}_0 = \{\sigma_1, \ldots, \sigma_s\} \subset \mathcal{F}$ be a minimal generating set. By the minimality of $\mathcal{F}_0$, for each $1 \leq i \leq s$ we have $\mathcal{F} = \mathcal{I}(\mathcal{F}_0) \subsetneq \mathcal{I}(\mathcal{F}_0 \setminus \{\sigma_i\})$ because $\mathcal{I}(\mathcal{F}_0) \subseteq \mathcal{I}(\mathcal{F}_0 \setminus \{\sigma_i\})$ but $\mathcal{F} \neq \mathcal{I}(\mathcal{F}_0 \setminus \{\sigma_i\})$. Therefore, for each $1 \leq i \leq s$ we can find some permutation $\tau_i \in S_n$ such that $\tau_i \in \mathcal{I}(\mathcal{F}_0 \setminus \{\sigma_i\}) \setminus \mathcal{F}$, and thus, $|\sigma_j \cap \tau_i| < t$ for all $i \neq j$. On the other hand, because $\tau_i \notin \mathcal{F} = \mathcal{I}(\mathcal{F}_0)$, we must have that $|\tau_i \cap \sigma_i| \geq t$ for all i.

We can think of families of permutations in S_n as n-uniform hypergraphs on $[n] \times [n]$ by assigning a set H_π of n ordered pairs to each permutation $\pi \in S_n$ as follows

$$H_\pi = \{(1, \pi(1)), \ldots, (n, \pi(n))\}.$$

Observe that for any two permutations π and π', $|H_\pi \cap H_{\pi'}| = |\pi \cap \pi'|$. We require this t-intersecting version of the Bollobás Set-Pairs Inequality, proven by Füredi [36] to bound $|\mathcal{F}_0|$:

Theorem 2.3.10. *Let $A_1, \ldots, A_m$ be sets of size a and $B_1, \ldots, B_m$ be sets of size b such that we have $|A_i \cap B_i| < t$ and $|A_i \cap B_j| \geq t$ for $1 \leq i < j \leq m$. Then $m \leq \binom{a+b-2t+2}{a-t+1}$.*

We apply this to the sets $\{(A_i, B_i)\}_{i=1}^{2s}$, where for $1 \leq i \leq s$ we take $A_i = H_{\sigma_i}$ and $B_i = H_{\tau_i}$, and for $s+1 \leq i \leq 2s$ we set $A_i = H_{\tau_{i-s}}$ and $B_i = H_{\sigma_{i-s}}$. Hence, from the discussion above for $1 \leq i, j \leq s$ we have that $\left|H_{\sigma_i} \cap H_{\tau_j}\right| = |\sigma_i \cap \tau_j| < t$ if and only if $i = j$. Likewise for $s+1 \leq i, j \leq 2s$ we have that $\left|H_{\sigma_{i-s}} \cap H_{\tau_{j-s}}\right| = |\sigma_{i-s} \cap \tau_{j-s}| < t$ if and only if $i = j$. Because $\mathcal{F}$ is intersecting then so is $\mathcal{F}_0 \subseteq \mathcal{F}$ and $\left|H_{\sigma_i} \cap H_{\sigma_{j-s}}\right| \geq t$ for all $1 \leq i \leq s$ and $s+1 \leq j \leq 2s$ because $|\sigma_i \cap \sigma_{j-s}| \geq t$. Thus, the conditions of Theorem 2.3.10 are satisfied, and hence we deduce that $2s \leq \binom{2n-2t+2}{n-t+1}$. Thus, $|\mathcal{F}_0| \leq \frac{1}{2}\binom{2n-2t+2}{n-t+1}$.

Thus, to every maximal t-intersecting family $\mathcal{F}$ we may assign a distinct generating set of at most $\frac{1}{2}\binom{2n-2t+2}{n-t+1}$ permutations, giving the above sum as a bound on the number of maximal t-intersecting families. The upper bound follows since $n! \leq n^n$ and $\binom{2n-2t+2}{n-t+1} \leq 2^{2n-2t+2}$. $\square$

Given this bound, we apply Lemmas 2.2.3 and 2.2.5 to prove our enumerative results. Proposition 2.3.9 shows that we may take $M = n^{n2^{2n-2t+1}}$. Each trivial t-intersecting family, on the other hand, has to fix the images of t indices. There are $\binom{n}{t}$ ways to choose the indices, $\binom{n}{t}$ ways to choose their images, and $t!$ ways to assign the images to the indices, and thus $T = \binom{n}{t}^2 t!$ maximal trivial t-intersecting families. The required extremal result is due to Ellis, Friedgut, and Pilpel [27], who showed that for n sufficiently large with respect to t, the largest t-intersecting families in S_n are the trivial ones, with size $N_0 = (n-t)!$. Moreover, note that there are at least $t+1$ fixed indices in the intersection of two trivial t-intersecting families, and so two distinct maximal t-intersecting families can have at most $N_2 = (n-t-1)!$ members in common. The stability result obtained by Ellis [25] shows that when t is fixed and n tends to infinity, the largest non-trivial t-intersecting family has size $N_1 = (1 - 1/e + o(1))\,(n-t)!$. Combining these ingredients, we now prove Theorems 2.3.6 and 2.3.7.

Proof of Theorem 2.3.6. We first apply Lemma 2.2.3 to show that almost all t-intersecting families are trivial. We have

$$\log M + N_1 - N_0 = n2^{2n-2t+1}\log n - (1/e + o(1))\,(n-t)! \to -\infty,$$

and so (2.1) is satisfied. This shows that the number of non-trivial t-intersecting families is $o\left(2^{(n-t)!}\right)$.

We use Lemma 2.2.5 to count the number of trivial t-intersecting families. We see that (2.3) holds, since

$$2\log T + N_2 - N_0 = 2\log\left(\binom{n}{t}^2 t!\right) + (n-t-1)! - (n-t)! \le 4t\log(nt) - (n-t-1)(n-t-1)! \to -\infty.$$

Hence the number of trivial t-intersecting families is $\left(\binom{n}{t}^2 t! + o(1)\right)2^{(n-t)!}$. As the non-trivial t-intersecting families constitute a lower-order term, this completes the proof. □

Proof of Theorem 2.3.7. To prove that almost every t-intersecting family of m permutations is trivial, we show that (2.2) is satisfied. Indeed, for $m \ge n2^{2n-2t+2}\log n$,

$$\begin{aligned}\log M - m\log\left(\frac{N_0}{N_1}\right) &= n2^{2n-2t+1}\log n - m\log\left(\frac{(n-t)!}{(1-1/e+o(1))\,(n-t)!}\right)\\ &\le n2^{2n-2t+1}\log n - 0.6m \to -\infty.\end{aligned}$$

□

Finally, we seek to prove Theorem 2.3.8, showing that when $p \ge \frac{800n2^{2n-2t}\log n}{(n-t)!}$, with high probability the largest t-intersecting family in the p-random set of permutations $(S_n)_p$ is trivial. Let $\mathcal{T} \subset S_n$ be a fixed maximal trivial t-intersecting family, and let $\mathcal{F}_1, \dots, \mathcal{F}_M$ be all of the maximal non-trivial t-intersecting families. Then the largest trivial t-intersecting family in $(S_n)_p$ has size at least $|(\mathcal{T})_p|$, while the largest non-trivial t-intersecting family has size $\max_i |(\mathcal{F}_i)_p|$. In expectation, $\mathbb{E}\left[|(\mathcal{T})_p|\right] = p\,|\mathcal{T}| > p\,|\mathcal{F}_i| = \mathbb{E}\left[|(\mathcal{F}_i)_p|\right]$, and our bound on M is strong enough for a union bound calculation to go through. We require the following version of Hoeffding's Inequality that is derived from [59, Theorem 2.3].

Theorem 2.3.11 (Hoeffding). *Let the random variables $X_1, X_2, \dots, X_n$ be independent, with $0 \le X_k \le 1$ for each k. Let $X = \sum_{k=1}^n X_k$, let $\mu = \mathbb{E}[X]$. Then, for any $\varepsilon > 0$,*

$$\mathbb{P}\left(X \ge (1+\varepsilon)\mu\right) \le \exp\left(-\frac{1}{2}\varepsilon^2\mu\right) \quad \text{and} \quad \mathbb{P}\left(X \le (1-\varepsilon)\mu\right) \le \exp\left(-\frac{1}{2}\varepsilon^2\mu\right).$$

Proof of Theorem 2.3.8. Let $(\mathcal{T})_p = \mathcal{T} \cap (S_n)_p$, let $(\mathcal{F}_i)_p = \mathcal{F}_i \cap (S_n)_p$, and set $\varepsilon = 1/10$. Let E_0 be the event that $|(\mathcal{T})_p| < (1-\varepsilon)p|\mathcal{T}| = (1-\varepsilon)pN_0$, and let E_i be the event that $|(\mathcal{F}_i)_p| > (1+\varepsilon)pN_1$. Since $N_0 = (n-t)!$ and $N_1 = (1 - 1/e + o(1))\,(n-t)!$, we have $(1+\varepsilon)pN_1 < (1-\varepsilon)pN_0$. If there is a non-trivial largest t-intersecting family in $(S_n)_p$, we must have $\max_i |(\mathcal{F}_i)_p| \ge |(\mathcal{T})_p|$, and so at least one of the events E_j, $0 \le j \le M$, must hold.

Now $|(\mathcal{T})_p| \sim \text{Bin}(N_0, p)$, and so applying Theorem 2.3.11 with $\mu = pN_0$, we have $\mathbb{P}(E_0) \le \exp\left(-\frac{pN_0}{200}\right)$. Similarly, for $1 \le i \le M$, $|(\mathcal{F}_i)_p| \sim \text{Bin}(|\mathcal{F}_i|, p)$, where $|\mathcal{F}_i| \le N_1$. Let $X \sim \text{Bin}(N_1, p)$. Applying Theorem 2.3.11 to X with $\mu = pN_1$, we have

$$\mathbb{P}(E_i) = \mathbb{P}(|(\mathcal{F}_i)_p| \ge (1+\varepsilon)pN_1) \le \mathbb{P}(X \ge (1+\varepsilon)pN_1) \le \exp\left(-\frac{pN_1}{200}\right).$$

Hence, by the union bound,

$$\mathbb{P}\left(\bigcup_{i=0}^{M} E_i\right) = \exp\left(-\frac{pN_0}{200}\right) + M\exp\left(-\frac{pN_1}{200}\right) \le \left(n^{n2^{2n-2t+1}} + 1\right) \cdot \exp\left(-\frac{pN_1}{200}\right) = o(1)$$

when $p \ge \frac{800n2^{2n-2t}\log n}{(n-t)!} \ge \frac{200}{N_1} n2^{2n-2t+1}\log n$. Thus, for such p, the largest t-intersecting families in $(S_n)_p$ are trivial with high probability. □

2.4 Hypergraphs

For $k \ge 2$ and $1 \le t < k$, a k-uniform hypergraph $\mathcal{H}$ on vertex set $[n]$ is *t-intersecting* if every pair of edges shares at least t vertices. As before, in the case when $t = 1$, we call such families *intersecting*. A t-intersecting family $\mathcal{H}$ is said to be *trivial* if every edge in $\mathcal{H}$ contains a fixed set of t vertices. In a classical paper from 1961, Erdős, Ko, and Rado [30] proved the following analogue of Theorem 2.1.3 for t-intersecting k-uniform hypergraphs on $[n]$.

Theorem 2.4.1. *For $n \ge t + (k-t)\binom{k}{t}^3$, the largest t-intersecting k-uniform hypergraphs on $[n]$ have size at most $\binom{n-t}{k-t}$.*

The following generalization of Theorem 2.4.1 was shown by Frankl [33] in 1978 for $t > 14$ and completed for all values of t in a result by Wilson [73] in 1984:

Theorem 2.4.2. *For $n \geq (t+1)(k-t+1)$, the largest t-intersecting k-uniform hypergraphs on $[n]$ have size at most $\binom{n-t}{k-t}$.*

Note that $n \geq (t+1)(k-t+1)$ is the best possible strengthening of the original result, Theorem 2.4.1, and for intersecting families, when $t = 1$, this gives Theorem 2.1.3 when $n \geq 2k$.

We show just beyond this bound that additionally the trivial t-intersecting k-uniform hypergraphs on $[n]$ are actually typical.

Theorem 2.4.3. *Let n, $k = k(n) \geq 3$ and $t = t(n) \geq 1$ be integers such that $n \geq (t+1)(k-t+1) + \eta_{k,t}$, where*

$$\eta_{k,t} = \begin{cases} k + 8\ln k & \text{for } t = 1, \\ 12\ln k & \text{for } t = 2 \text{ and } k - t \geq 3, \\ 1 & \text{for } t \geq 3 \text{ and } k - t \geq 3 \\ 31 & \text{for } t \geq 2 \text{ and } k - t = 2, \\ 18k & \text{for } t \geq 2 \text{ and } k - t = 1. \end{cases}$$

Almost all t-intersecting k-uniform hypergraphs on $[n]$ are trivial, and there are $\left(\binom{n}{t} + o(1)\right) 2^{\binom{n-t}{k-t}}$ t-intersecting k-uniform hypergraphs.

Observe that for $\eta_{k,t} = 1$, which we have for most values of t and k, this is the best possible result; when $n = (t+1)(k-t+1)$ the largest non-trivial t-intersecting hypergraphs on $[n]$ are as large as the trivial t-intersecting hypergraphs on $[n]$. In fact, there are many more of them, and hence for this n almost every t-intersecting hypergraph on $[n]$ is non-trivial.

However, there is no doubt that the case when $t = 1$ is the most natural and interesting to study. Theorem 2.4.3 gives the asymptotic number of intersecting hypergraphs when $n \geq 3k + 8\ln k$. On the other hand, it is known that the trivial hypergraphs are the largest when $n \geq 2k$, and uniquely so when $n \geq 2k+1$. The following theorem, which we prove using spectral methods and the theory of graph containers, provides a slightly weaker result that covers the entire range. The proof, along with an exposition of the graph containers method, appears in Section 2.6.

Theorem 2.6.1. *For $k \geq 3$ and $n \geq 2k+1$, let $I(n,k)$ denote the number of intersecting k-uniform hypergraphs on $[n]$. Then*

$$\log I(n,k) = (1 + o(1)) \binom{n-1}{k-1}.$$

Similarly to permutations, we are able to obtain a sparse version of the Erdős–Ko–Rado Theorem. Let $\mathcal{H}^k(n,p)$ denote the p-random k-uniform hypergraph on $[n]$, in which every edge in $\binom{[n]}{k}$ is included independently with probability p. Balogh, Bohman, and Mubayi [7] initiated the study of intersecting hypergraphs in the sparse random setting; among other results, they determined the size of the largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ when $k < n^{1/2-\varepsilon}$. Recently, Gauy, Hàn, and Oliveira [37] determined the asymptotic size of the largest intersecting family for all k and almost all p. Hamm and Kahn [46] obtained an exact result for $k < (\frac{1}{4}-c)(n\log n)^{1/2}$ for some small constant c and $p \gg \binom{n-k}{k}^{-1}$, showing that with high probability every largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ is trivial. We prove that the same holds for k as large as linear in n, provided p is somewhat larger. Independently Hamm and Kahn [40] studied the case $n = 2k+1$ and $p = 1-c$ for some constant $c > 0$.

Theorem 2.4.4. *For $3 \le k \le \frac{n}{4}$, if*

$$p = p(n) \ge \frac{9n\log\left(\frac{ne}{k}\right)\binom{2k}{k}\binom{n}{k}}{\binom{n-k}{k}^2}, \tag{2.4}$$

then with high probability every largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ is trivial.

Observe that the lower bound on p in (2.4) is at most $9n\log\left(\frac{ne}{k}\right)\left(\frac{2kn}{(n-k)^2}\right)^k$, and is thus exponentially small with respect to $k\log\left(\frac{n}{k}\right)$.

We now return our attention to proving Theorems 2.4.3 and 2.4.4. The proof of Theorem 2.6.1 uses a different method, and will be given in Section 2.6. We begin with a bound on the number of maximal t-intersecting hypergraphs:

Proposition 2.4.5. *The number of maximal t-intersecting k-uniform hypergraphs on $[n]$ is at most*

$$\sum_{i=1}^{\binom{2(k-t)+1}{k-t}} \binom{\binom{n}{k}}{i} \le \binom{n}{k}^{\binom{2(k-t)+1}{k-t}}.$$

Proof. The proof of this proposition follows the proof of Proposition 2.2.2, except we must replace Theorem 2.2.1 with its t-intersecting version Theorem 2.3.10. This shows that every maximal t-intersecting hypergraph yields a minimal generating set of at most $\frac{1}{2}\binom{2(k-t)+2}{k-t+1} = \binom{2(k-t)+1}{k-t}$ edges. Similar to before, the number of maximal t-intersecting hypergraphs is bounded above by the number of sets of at most $\binom{2(k-t)+1}{k-t}$ edges, giving the resulting upper bound above. $\square$

We shall now use Lemma 2.2.3 to show that almost every t-intersecting hypergraph is trivial. Proposition 2.4.5 supplies us with the value of M required. The Erdős–Ko–Rado theorem [30] states that for n sufficiently large, the largest t-intersecting hypergraphs are the trivial ones, which have size $N_0 = \binom{n-t}{k-t}$. Frankl [33] for $t > 14$ showed that $n \ge (t+1)(k-t+1)$ was the correct bound and Wilson [73] extended this result to all values of t.

Stability results for the Erdős–Ko–Rado Theorem have a long history, beginning with the Hilton–Milner Theorem [41], which resolved the $t = 1$ case. After much incremental progress, Ahlswede and Khachatrian [5] completely determined the largest non-trivial intersecting hypergraphs for all ranges of parameters. In our range of interest, $n \geq (t+1)(k-t+1)$, there are two possible largest non-trivial hypergraphs:

$$\mathcal{H}_1 = \{F : |F \cap [t+2]| \geq t+1\}, \text{ and}$$
$$\mathcal{H}_2 = \{F : [t] \subset F, F \cap [t+1, k+1] \neq \emptyset\} \cup \{[k+1] \setminus \{i\} : 1 \leq i \leq t\}.$$

Theorem 2.4.6 (Ahlswede–Khachatrian)**.** *Suppose* $n \geq (t+1)(k-t+1)$*. If* $k \leq 2t+1$*, then the largest non-trivial* t*-intersecting* k*-uniform hypergraph over* $[n]$ *has size* $|\mathcal{H}_1|$*. If* $k \geq 2t+2$*, then the largest non-trivial hypergraph has size* $\max\{|\mathcal{H}_1|, |\mathcal{H}_2|\}$*.*

This theorem provides the value of N_1 needed for Lemma 2.2.3. Before we proceed, we evaluate $|\mathcal{H}_1|$ and $|\mathcal{H}_2|$, making use of Pascal's identity for binomial coefficients.

$$|\mathcal{H}_1| = (t+2)\binom{n-t-2}{k-t-1} + \binom{n-t-2}{k-t-2} = \binom{n-t}{k-t} - \left(1 - \frac{(t+1)(k-t)}{n-t-1}\right)\binom{n-t-1}{k-t}. \tag{2.5}$$
$$|\mathcal{H}_2| = \binom{n-t}{k-t} - \binom{n-k-1}{k-t} + t. \tag{2.6}$$

In light of Theorem 2.4.6, we have $N_1 \leq \max\{|\mathcal{H}_1|, |\mathcal{H}_2|\}$, which we estimate by

$$\begin{aligned} N_1 &\leq \max\{|\mathcal{H}_1|, |\mathcal{H}_2|\} \\ &= \binom{n-t}{k-t} - \min\left\{\left(1 - \frac{(t+1)(k-t)}{n-t-1}\right)\binom{n-t-1}{k-t}, \binom{n-k-1}{k-t} - t\right\} \\ &\leq \binom{n-t}{k-t} - \left(1 - \frac{(t+1)(k-t)}{n-t-1}\right)\binom{n-k-1}{k-t} + t \leq \binom{n-t}{k-t} - \frac{1}{n}\binom{n-k-1}{k-t} + n, \end{aligned} \tag{2.7}$$

where the last inequality holds for $n \geq (t+1)(k-t+1) + 1$. We shall also use the following inequality for $a \geq b \geq r$:

$$\frac{\binom{a}{r}}{\binom{b}{r}} = \prod_{j=0}^{r-1} \frac{a-j}{b-j} \geq \left(\frac{a}{b}\right)^r. \tag{2.8}$$

Finally, to count the number of trivial families, we use Lemma 2.2.5. Since each trivial family fixes t elements, there are $T = \binom{n}{t}$ maximal trivial families. The intersection of any two such families must fix at least $t+1$ elements, and so can have size at most $N_2 = \binom{n-t-1}{k-t-1}$. With these preliminaries in place, we now prove Theorem 2.4.3.

Proof of Theorem 2.4.3. We will first prove that, for n, k and t as in the statement of the theorem, almost all t-intersecting hypergraphs are trivial. To this end, we verify that (2.1) of Lemma 2.2.3 holds.

We start with the case $t = 1$. The Hilton–Milner Theorem states that when $n > 2k$, the largest non-trivial intersecting hypergraph is $\mathcal{H}_2$, and so $N_1 = |\mathcal{H}_2| = \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$. Recall that the trivial hypergraphs have size $N_0 = \binom{n-1}{k-1}$. Finally, since $\binom{n}{k} \le 2^n$, Proposition 2.4.5 shows that we may use $\log M \le \binom{2k-1}{k-1} n$.

Hence, using (2.6) and (2.8), we have

$$\log M + N_1 - N_0 \le \binom{2k-1}{k-1} n - \binom{n-k-1}{k-1} + 1 \le \left(n - \left(\frac{n-k-1}{2k-1} \right)^{k-1} \right) \binom{2k-1}{k-1} + 1.$$

For $t = 1$, we have $n \ge (t+1)(k-t+1) + \eta_{k,t} = 2k + \eta_{k,1} = 3k + 8 \ln k$. We may bound

$$\left(\frac{n-k-1}{2k-1} \right)^{k-1} = \left(\frac{n-k-1}{2k-1} \right)^2 \left(\frac{n-k-1}{2k-1} \right)^{k-3} \ge \frac{n^2}{16k^2} \left(1 + \frac{8 \ln k}{2k-1} \right)^{k-3}.$$

Since $1 + x \ge \exp(6x/11)$ for $x \le 1$, when k is large we have

$$k^{-2} \left(1 + \frac{8 \ln k}{2k-1} \right)^{k-3} \ge k^{-2} \exp \left(\frac{48(k-3) \ln k}{22k} \right) \ge k^{-2} \exp(2 \ln k) = 1.$$

Thus there is some constant $c > 0$ such that $k^{-2} \left(1 + \frac{8 \ln k}{2k-1} \right)^{k-3} \ge c$ for all k, and thus $\left(\frac{n-k-1}{2k-1} \right)^{k-1} = \Omega(n^2)$. Hence it follows that $\left(n - \left(\frac{n-k-1}{2k-1} \right)^{k-1} \right) \binom{2k-1}{k-1} - 1 \to -\infty$.

We next handle the case $k - t = 1$. In this setting, we have $k \le 2t + 1$, and hence by Theorem 2.4.6, the largest non-trivial hypergraph has size $N_1 = |\mathcal{H}_1|$. Using $\binom{n}{k} \le \left(\frac{ne}{k} \right)^k$, we may use $\log M \le k \log \left(\frac{ne}{k} \right) \binom{2(k-t)+1}{k-t}$. Using (2.5) gives

$$\begin{aligned} \log M + N_1 - N_0 &\le k \log \left(\frac{ne}{k} \right) \binom{2(k-t)+1}{k-t} - \left(1 - \frac{(t+1)(k-t)}{n-t-1} \right) \binom{n-t-1}{k-t} \\ &= 3k \log \left(\frac{ne}{k} \right) - (n - 2k) = 3k \log \left(\frac{ne}{k} \right) + 2k - n. \end{aligned}$$

This expression is increasing in k. Since we are assuming $n \ge (t+1)(k-t+1) + \eta_{k,t} = 2k + \eta_{k,k-1} = 20k$, we substitute $k = n/20$ to obtain $\log M + N_1 - N_0 \le (3 \log(20e) - 18)\, n/20 \to -\infty$, since $3 \log(20e) < 18$.

Similar calculations show that when $t \ge 2$ and $k - t = 2$, $\eta_{k,k-2} = 31$ suffices. In this setting, we still have $N_1 = |\mathcal{H}_1|$. Using $\log M \le n \binom{2(k-t)+1}{k-t}$ and $n \ge (t+1)(k-t+1) + \eta_{k,k-2} > 3k$,

$$\log M + N_1 - N_0 \le (10 - \eta_{k,k-2}/3)\, n \to -\infty.$$

We now consider the remaining cases, when $t \geq 2$ and $k-t \geq 3$. In this range, the largest non-trivial hypergraph has size $N_1 = \max\{|\mathcal{H}_1|, |\mathcal{H}_2|\}$. Using $\binom{n}{k} \leq 2^n$, we have $\log M \leq n\binom{2(k-t)+1}{k-t} \leq 2n\binom{2(k-t)}{k-t}$. By (2.7) and (2.8), and observing that $n-k-1 \geq t(k-t)+\eta_{k,t}$, we have

$$\begin{aligned}\log M + N_1 - N_0 &\leq 2n\binom{2(k-t)}{k-t} - \frac{1}{n}\binom{n-k-1}{k-t} + n \\ &\leq \left(3n - \frac{1}{n}\left(\frac{n-k-1}{2(k-t)}\right)^{k-t}\right)\binom{2(k-t)}{k-t} \\ &\leq \left(3n - \frac{n^2}{64(k-t)^3}\left(\frac{t(k-t)+\eta_{k,t}}{2(k-t)}\right)^{k-t-3}\right)\binom{2(k-t)}{k-t}. \qquad (2.9)\end{aligned}$$

If $t = 2$, then $\eta_{k,t} = 12\ln k$, and $\frac{t(k-t)+\eta_{k,t}}{2(k-t)} = 1 + \frac{6\ln k}{k-2}$. Using $1 + x \geq \exp(6x/11)$ again, we find that for large k,

$$(k-2)^{-3}\left(1+\frac{6\ln k}{k-2}\right)^{k-5} \geq (k-2)^{-3}\exp\left(\frac{36(k-5)\ln k}{11k}\right) \geq k^{-3}\exp(3\ln k) = 1.$$

It follows that there is some constant $c > 0$ such that $(k-2)^{-3}\left(\frac{2(k-2)+\eta_{k,2}}{2(k-2)}\right)^{k-5} \geq c$ for all k.

If instead $t \geq 3$, then $(k-t)^{-3}\left(\frac{t(k-t)+\eta_{k,t}}{2(k-t)}\right)^{k-t-3} > (k-t)^{-3}(\frac{3}{2})^{k-t-3} \to \infty$ as $k-t \to \infty$, and thus there is some $c > 0$ such that $(k-t)^{-3}\left(\frac{t(k-t)+\eta_{k,t}}{2(k-t)}\right)^{k-t-3} \geq c$ for all $k > t$. Hence, in either case, $\frac{n^2}{64(k-t)^3}\left(\frac{t(k-t)+\eta_{k,t}}{2(k-t)}\right)^{k-t-3} = \Omega(n^2)$, and so from (2.9) it follows that $\log M + N_1 - N_0 \to -\infty$.

Thus our choice of $\eta_{k,t}$ ensures that for all $k > t$ we have $\log M + N_1 - N_0 \to -\infty$, satisfying (2.1) of Lemma 2.2.3, thus showing that almost all t-intersecting k-uniform hypergraphs are trivial. To complete the proof of Theorem 2.4.3, we need only count the number of trivial hypergraphs. By Lemma 2.2.5, it suffices to verify (2.3). We have

$$2\log T + N_2 - N_0 = 2\log\binom{n}{t} + \binom{n-t-1}{k-t-1} - \binom{n-t}{k-t} \leq 2t\log\left(\frac{ne}{t}\right) - \binom{n-t-1}{k-t} \to -\infty$$

for $k-t \geq 2$ or $k-t=1$ and $n \geq 20t$. It follows that there are $\left(\binom{n}{t} + o(1)\right)2^{\binom{n-t}{k-t}}$ t-intersecting k-uniform hypergraphs on $[n]$, as claimed. □

We conclude this section with the proof of Theorem 2.4.4, showing that even in sparse random hypergraphs, when the edge probability is as given in (2.4) the largest intersecting subhypergraphs are trivial.

Proof of Theorem 2.4.4. The proof follows that of Theorem 2.3.8. Let $\mathcal{T}$ denote a fixed maximal trivial hypergraph, and let $(\mathcal{T})_p = \mathcal{T} \cap \mathcal{H}^k(n,p)$ be those edges of $\mathcal{T}$ selected in $\mathcal{H}^k(n,p)$. Let $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_M$ be the maximal non-trivial hypergraphs, where by Proposition 2.4.5 we have $M < \binom{n}{k}^{\binom{2k-1}{k-1}} < 2^{k\log\left(\frac{ne}{k}\right)\binom{2k-1}{k-1}}$, and let $(\mathcal{F}_i)_p = \mathcal{F}_i \cap \mathcal{H}^k(n,p)$ denote the corresponding random subhypergraphs.

Observe that $|\mathcal{T}| = N_0 = \binom{n-1}{k-1}$, while by the Hilton–Milner Theorem [41], $|\mathcal{F}_i| \leq N_1 = \binom{n-1}{k-1} - \binom{n-k-1}{k-1} + 1$. Setting $\tau = p\binom{n-k-1}{k-1}/3$, define events $E_0 = \{|(\mathcal{T})_p| \leq pN_0 - \tau\}$ and $E_i = \{|(\mathcal{F}_i)_p| \geq pN_1 + \tau\}$ for $1 \leq i \leq M$. By our choice of τ, if none of the events $\{E_i\}_{i=0}^M$ occur then $|(\mathcal{T})_p| > \max_i \{|(\mathcal{F}_i)_p|\}$, and so the largest intersecting subhypergraphs in $\mathcal{H}^k(n,p)$ are trivial.

Applying Theorem 2.3.11, we find

$$\mathbb{P}(E_0) \leq \exp\left(-\frac{\tau^2}{2pN_0}\right) \quad \text{and} \quad \mathbb{P}(E_i) \leq \exp\left(-\frac{\tau^2}{2pN_1}\right) \leq \exp\left(-\frac{\tau^2}{2pN_0}\right).$$

Hence, by the union bound,

$$\mathbb{P}\left(\cup_{i=0}^M E_i\right) \leq (M+1)\exp\left(-\frac{\tau^2}{2pN_0}\right) \leq \left(2^{k\log\left(\frac{ne}{k}\right)\binom{2k-1}{k-1}} + 1\right)\exp\left(-\frac{p\binom{n-k-1}{k-1}^2}{18\binom{n-1}{k-1}}\right) \to 0$$

when $p \geq \frac{9n\log\left(\frac{ne}{k}\right)\binom{2k}{k}\binom{n}{k}}{\binom{n-k}{k}^2} \geq \frac{18k\log\left(\frac{ne}{k}\right)\binom{2k-1}{k-1}\binom{n-1}{k-1}}{\binom{n-k-1}{k-1}^2}$, giving the bound in (2.4). $\square$

As proven in [7], when $k \gg \sqrt{n\log\log n}$ and $\frac{\log n}{\binom{n-1}{k}} \ll p \ll \frac{e^{k^2/2n}}{\binom{n}{k}}$, a simple first moment argument shows that the largest intersecting subhypergraph of $\mathcal{H}^k(n,p)$ is non-trivial with high probability. This holds for p considerably smaller than in (2.4), and it would be very interesting to determine the threshold at which trivial hypergraphs become the largest intersecting subhypergraphs of $\mathcal{H}^k(n,p)$.

2.5 Vector Spaces

Let V be an n-dimensional vector space over a finite field $\mathbb{F}_q$. The number of k-dimensional subspaces in V 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 V is *intersecting* if $\dim(F_1 \cap F_2) \geq 1$ for all pairs of subspaces $F_1, F_2 \in \mathcal{F}$. Taking all k-dimensional subspaces through a given 1-dimensional subspace yields a *trivial* intersecting family in this context. Hsieh [42] proved an Erdős–Ko–Rado-type Theorem for vector spaces:

Theorem 2.5.1. *For $n \geq 2k+1$, any intersecting family $\mathcal{F}$ of k-dimensional subspaces of V has size at most $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$; and equality holds only if $\mathcal{F}$ is trivial.*

The results we obtain for permutations and hypergraphs can be extended to vector spaces as well, and we determine here the typical structure of intersecting families of subspaces.

Theorem 2.5.2. *If $k \geq 2$, and either $q = 2$ and $n \geq 2k+2$ or $q \geq 3$ and $n \geq 2k+1$, almost all intersecting families of k-dimensional subspaces of $\mathbb{F}_q^n$ are trivial, and there are $\left(\begin{bmatrix} n \\ 1 \end{bmatrix}_q + o(1) \right) 2^{\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q}$ intersecting families.*

This result is related to algebraic geometry as the Grassmannian is the set of all k-dimensional subspaces in an n-dimensional vector space.

In this section we prove Theorem 2.5.2, showing that almost all intersecting families of subspaces of a finite vector space are trivial. We begin, as always, with a bound on the number of maximal families.

Proposition 2.5.3. *The number of maximal intersecting families of k-dimensional subspaces of $\mathbb{F}_q^n$ is at most*

$$\sum_{i=0}^{\binom{2k-1}{k-1}} \binom{\begin{bmatrix} n \\ k \end{bmatrix}_q}{i} \leq \begin{bmatrix} n \\ k \end{bmatrix}_q^{\binom{2k-1}{k-1}}.$$

Proof. Once again, we follow the strategy of Proposition 2.2.2, seeking to show that every maximal intersecting family of subspaces contains a minimal generating set of at most $\frac{1}{2}\binom{2k}{k} = \binom{2k-1}{k-1}$ subspaces. We must replace Theorem 2.2.1 with its vector space analogue, proven by Lovász [55] and appearing in the form below in [6].

Theorem 2.5.4 (Lovász)**.** *Let $U_1, \ldots, U_m$ be a-dimensional and $V_1, \ldots, V_m$ be b-dimensional subspaces of a vector space W over a field $\mathbb{F}$ such that $U_i \cap V_i = \{0\}$ and $U_i \cap V_j \neq \{0\}$ for $1 \leq i < j \leq m$. Then $m \leq \binom{a+b}{a}$.*

This gives a map from maximal intersecting families of subspaces to sets of at most $\binom{2k-1}{k-1}$ subspaces, resulting in the above bound. $\square$

This proposition gives a value for M to be used when applying Lemma 2.2.3. As stated above, the corresponding extremal result was proven by Hsieh [42], who showed that when $n \geq 2k+1$, the largest intersecting families are trivial, with size $N_0 = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$. Each trivial intersecting family fixes a one-dimensional subspace, and hence there are $T = \begin{bmatrix} n \\ 1 \end{bmatrix}_q$ maximal trivial intersecting families. The intersection of any two fixes a two-dimensional subspace, and thus has size $N_2 = \begin{bmatrix} n-2 \\ k-2 \end{bmatrix}_q$. A stability result was obtained by Blokhuis, Brouwer, Chowdhury, Frankl, Mussche, Patkós, and Szőnyi [13]:

Theorem 2.5.5. *For $q \geq 3$ and $n \geq 2k+1$ or $q = 2$ and $n \geq 2k+2$, the size of a largest non-trivial intersecting family is $\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix} + q^k$.*

We set $N_1 = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix} + q^k$ when $q \geq 3$ and $n \geq 2k+1$ or $q = 2$ and $n \geq 2k+2$. With these results in hand, we prove Theorem 2.5.2.

Proof of Theorem 2.5.2. We shall first verify that (2.1) of Lemma 2.2.3 holds, thus showing that almost all intersecting families are trivial. Since either $q \geq 3$ and $n \geq 2k+1$ or $q = 2$ and $n \geq 2k+2$, the extremal and stability results hold, and thus $M \leq \begin{bmatrix} n \\ k \end{bmatrix}_q^{\binom{2k-1}{k-1}}$, $N_0 = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q$ and $N_1 = \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix}_q + q^k$. This gives

$$\log M + N_1 - N_0 = \log \left(\begin{bmatrix} n \\ k \end{bmatrix}_q \right) \binom{2k-1}{k-1} + q^k - q^{k(k-1)} \begin{bmatrix} n-k-1 \\ k-1 \end{bmatrix}_q . \tag{2.10}$$

We bound the Gaussian binomial coefficients above and below by

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

and use the fact that $\binom{2k-1}{k-1} < 4^k$ to show that the right-hand side of (2.10) is at most

$$k(n-k)4^k \log(2q) + q^k - q^{k(k-1)} \cdot q^{(n-2k)(k-1)} \leq n^2 4^k \log(2q) + q^k - q^{(n-k)(k-1)}. \tag{2.11}$$

If $k = 2$, then the right-hand side of (2.11) is $16n^2 \log(2q) + q^2 - q^{n-2} \to -\infty$ as $n \to \infty$. On the other hand, if $3 \leq k < n/2$ then $(n-k)(k-1) \geq 2(n-2)$, so the right-hand side of (2.11) is bounded above by $n^2 2^n \log(2q) + q^{n/2} - q^{2(n-2)} \to \infty$, since $q^2 > 2$. In either case, we have $\log M + N_1 - N_0 \to -\infty$, and, by Lemma 2.2.3, almost all intersecting families are trivial.

Now we need only show that there are $\left(\begin{bmatrix} n \\ 1 \end{bmatrix}_q + o(1) \right) 2^{\begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q}$ trivial families, which will follow by verifying (2.3) and applying Lemma 2.2.5. We have

$$\begin{aligned} 2\log T + N_2 - N_0 &= 2\log\left(\begin{bmatrix} n \\ 1 \end{bmatrix}_q \right) + \begin{bmatrix} n-2 \\ k-2 \end{bmatrix}_q - \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \\ &\leq 2n \log q - \left(1 - \frac{q^{k-1}-1}{q^{n-1}-1} \right) \begin{bmatrix} n-1 \\ k-1 \end{bmatrix}_q \leq 2n \log q - \frac{1}{2} q^{k(n-k)} \to -\infty, \end{aligned}$$

as required. This completes the proof. □

2.6 Exposition with Containers

Classical problems from a number of different areas can be reformulated into questions concerning independent sets in carefully constructed auxiliary hypergraphs. For intersecting families of k-element subsets of the ground set $[n] := \{1, 2, \ldots, n\}$, for example, the appropriate auxiliary graph has the vertex set consisting of all k-element subsets of $[n]$ with two vertices adjacent if and only if they are disjoint. Known as the Kneser graph $KG(n, k)$ in the literature, this graph has the property that subsets of vertex set correspond to k-uniform hypergraphs on $[n]$. Independent sets in the Kneser graph are of particular interest as they correspond directly to intersecting hypergraphs. Thus, the problem of counting intersecting k-uniform hypergraphs on $[n]$ reduces to counting the number of independent sets in $KG(n,k)$.

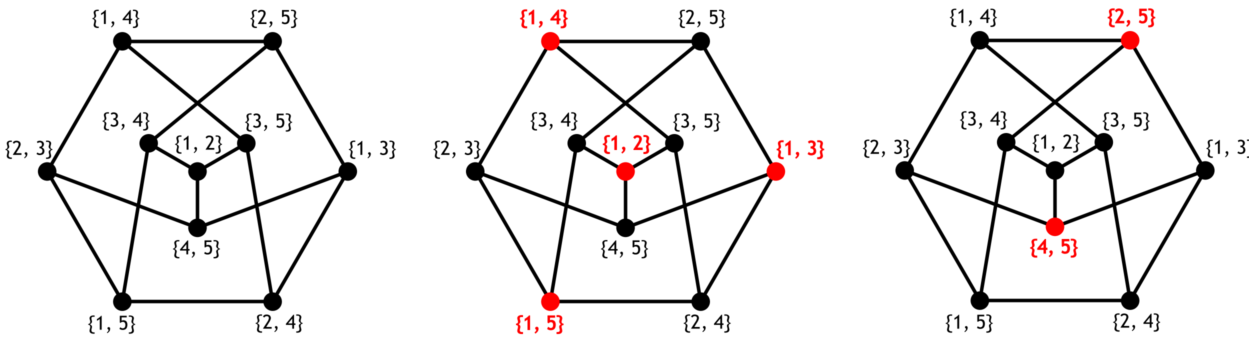

Figure 2.1: Independent Sets in Kneser Graphs

Independent sets in the Kneser graph, $KG(n,k)$, correspond to intersecting k-uniform hypergraphs on $[n]$. Note that the Petersen graph is isomorphic to the Kneser graph $KG(5,2)$.

Let $i(\mathcal{H})$ denote the number of independent sets in hypergraph $\mathcal{H}$. Every subset of a fixed independent set is also an independent set; this is true for a largest independent set in $\mathcal{H}$ so there must be at least $2^{\alpha(\mathcal{H})}$ independent sets where the independence number, denoted $\alpha(\mathcal{H})$, is the size of a largest independent set in $\mathcal{H}$. On the other hand, an upper bound for $i(\mathcal{H})$ can be obtained by enumerating all possible candidates for an independent set of $\mathcal{H}$, namely any possible subset of the vertex set of size at most $\alpha(\mathcal{H})$. Thus, trivial bounds for $i(\mathcal{H})$ are

$$2^{\alpha(\mathcal{H})} \leq i(\mathcal{H}) \leq \sum_{i=0}^{\alpha(\mathcal{H})} \binom{v(\mathcal{H})}{m}.$$

For hypergraphs meeting certain technical conditions Balogh, Morris, and Samotij and independently Saxton and Thomassen show the correct value of $i(\mathcal{H})$ is closer to the trivial lower bound as

$$2^{\alpha(\mathcal{H})} \leq i(\mathcal{H}) \leq 2^{(1+o(1))\alpha(\mathcal{H})}.$$

For a number of hypergraphs corresponding to nice problems, this so called "containers method" applies. For graphs, 2-uniform hypergraphs, the situation is more easily stated. Inspired by the work of Kleitman and Winston [47], this approach was first used by Alon, Balogh, Morris, and Samotij [1] to count sum-free subsets of abelian groups.

Although Theorem 2.4.3 provides very sharp results, it is somewhat incomplete in the case $t = 1$, as we require $n \geq 3k + 8\ln k$ instead of the Erdős–Ko–Rado threshold $n \geq 2k+1$. In this section we prove

Theorem 2.6.1, which fills in the gap with a slightly weaker result, providing the asymptotics of the *logarithm* of the number of intersecting hypergraphs. We prove the following result using graph containers.

Theorem 2.6.1. *For $k \geq 3$ and $n \geq 2k+1$, let $I(n,k)$ denote the number of intersecting k-uniform hypergraphs on $[n]$. Then*

$$\log I(n,k) = (1+o(1))\binom{n-1}{k-1}.$$

To prove this, we consider the following formulation of graph containers, appearing in the form below in a paper by Kohayakawa, Lee, Rödl, and Samotij [48].

Theorem 2.6.2. *Let G be a graph on N vertices, let R and ℓ be integers, and let $\beta > 0$ be a positive real. Then, provided*

$$e^{-\beta\ell} N \leq R, \tag{2.12}$$

and, for every subset $S \subset V(G)$ of at least R vertices, we have

$$e(S) \geq \beta\binom{|S|}{2}, \tag{2.13}$$

there is a collection of sets $C_i \subset V(G)$, $1 \leq i \leq \binom{N}{\ell}$, such that $|C_i| \leq R+\ell$ for every i and, for every independent set $I \subset V(G)$, there is some i satisfying $I \subset C_i$.

The condition of (2.13) in Theorem 2.6.2 requires large vertex subsets to induce subgraphs of positive density; this is referred to as supersaturation in the literature. For hypergraphs supersaturation is rather technical and can be difficult to show, but for graphs in particular this boils down to a checking a local density condition. If a graph G satisfies this condition, then the "containers method" shows that each independent set $\mathcal{I}$ in G can be labeled with a small subset $S \subset \mathcal{I}$ of size ℓ such that all independents sets labeled with $S \in \binom{[n]}{\ell}$ are essentially contained in a single set $C_i \subset V(G)$ that has few edges of G.

In the case at hand, namely the auxiliary graph in question is the Kneser graph $KG(n,k)$, supersaturation can be confirmed using spectral techniques; a similar approach is used by Gauy, Hàn, and Oliveira in [37]. A convenient way to relate the eigenvalues of a graph to the distribution of edges is the Expander Mixing Lemma, of Alon and Chung [2].

Theorem 2.6.3 (Expander Mixing Lemma)**.** *Let G be a D-regular graph on N vertices, and let λ be its second-largest eigenvalue in absolute value. Then for all $S \subseteq V(G)$,*

$$e(G[S]) \geq \frac{D}{2N}|S|^2 + \frac{\lambda}{2N}|S|\left(N - |S|\right).$$

By definition, the Kneser graph $KG(n, k)$ is a regular graph on $N = \binom{n}{k}$ vertices because there are $\binom{n}{k}$ ways to pick all possible k-element subsets from $[n]$. The degree of any vertex in $KG(n, k)$ is $D = \binom{n-k}{k}$ because the neighborhood of a vertex should consist of all k-element subsets that do not contain any elements of the k-element set of the original vertex. Fortunately all of the eigenvalues of the Kneser graph are well-known, they were determined in a seminal paper of Lovász [56].

Theorem 2.6.4. *The eigenvalues of the Kneser graph $KG(n, k)$ are*

$$(-1)^i \binom{n-k-i}{k-i},$$

where $i = 0, \ldots,\ k$.

Thus, the second largest eigenvalue (in absolute value) of the Kneser graph $KG(n, k)$ is

$$\lambda := -\binom{n-k-1}{k-1} = -\frac{k}{n-k}\binom{n-k}{k} = -\frac{k}{n-k}D.$$

Combined with Theorem 2.6.3, this gives the following supersaturation bound.

Proposition 2.6.5. *Given $\varepsilon > 0$, any set S of at least $(1+\varepsilon)\binom{n-1}{k-1}$ vertices in the Kneser graph $KG(n, k)$ induces at least $\left(1 - \frac{1}{1+\varepsilon}\right)\frac{Dn}{N(n-k)}\binom{|S|}{2}$ edges.*

Proof. Given a vertex set S with $|S| \geq (1+\varepsilon)\binom{n-1}{k-1} = (1+\varepsilon)\frac{kN}{n}$, we apply Theorem 2.6.3 and the fact that the second largest eigenvalue is $\lambda = -\frac{k}{n-k}D$ to obtain that

$$e(G[S]) \geq \frac{D}{2N}|S|^2 + \frac{\lambda}{2N}|S|\,(N - |S|) \geq \left(\frac{D-\lambda}{N} + \frac{\lambda}{|S|}\right)\binom{|S|}{2} \geq \left(1 - \frac{1}{1+\varepsilon}\right)\frac{Dn}{N(n-k)}\binom{|S|}{2}.$$

□

Having established supersaturation, we may now apply Theorem 2.6.2 to find a small set of containers of independent sets in the Kneser graph, from which we shall derive Theorem 2.6.1.

Proposition 2.6.6. *For $\varepsilon > 0$ and $2 \leq k \leq \frac{n-1}{2}$, let $R = (1+\varepsilon)\binom{n-1}{k-1}$ and*

$$\ell = \frac{1+\varepsilon}{\varepsilon} \cdot \frac{(n-k)\binom{n}{k}}{n\binom{n-k}{k}} \ln\left(\frac{n}{(1+\varepsilon)k}\right).$$

Then there exist k-uniform hypergraphs $\mathcal{F}_i$ over $[n]$, $1 \leq i \leq \binom{\binom{n}{k}}{\ell}$, each of size at most $R + \ell$, such that every intersecting k-uniform hypergraph $\mathcal{F}$ over $[n]$ is a subhypergraph of $\mathcal{F}_i$ for some i.

Proof. We apply Theorem 2.6.2 to the Kneser graph $KG(n,k)$. By Proposition 2.6.5, Condition (2.13) is satisfied by taking

$$\beta = \left(1 - \frac{1}{1+\varepsilon}\right) \frac{Dn}{N(n-k)},$$

where $D = \binom{n-k}{k}$ and $N = \binom{n}{k}$. In order to satisfy (2.12), we take

$$\ell = \frac{1}{\beta} \ln\left(\frac{N}{R}\right) = \frac{1}{\beta} \ln\left(\frac{n}{(1+\varepsilon)k}\right) = \frac{1+\varepsilon}{\varepsilon} \cdot \frac{(n-k)\binom{n}{k}}{n\binom{n-k}{k}} \ln\left(\frac{n}{(1+\varepsilon)k}\right).$$

Applying Theorem 2.6.2, the result follows by taking $\mathcal{F}_i$ to be the hypergraph with edges $C_i \subset \binom{[n]}{k}$, since every intersecting k-uniform hypergraph is an independent set of $KG(n,k)$. $\square$

From this we derive Theorem 2.6.1.

Proof of Theorem 2.6.1. Since there is an intersecting hypergraph of size $\binom{n-1}{k-1}$, and each of its subhypergraphs is also intersecting, we have a lower bound $\log I(n,k) \geq \binom{n-1}{k-1}$. We therefore need to show that $\log I(n,k) \leq (1+o(1))\binom{n-1}{k-1}$. Using Proposition 2.6.6, we will show that for any $\varepsilon > 0$, $\log I(n,k) \leq (1+2\varepsilon)\binom{n-1}{k-1}$, provided $n \geq 2k+1$ is sufficiently large with respect to ε.

We know that every intersecting hypergraph is contained in one of $\binom{N}{\ell}$ containers, each of size at most $R+\ell$, where R and ℓ are as in the statement of the proposition. By a simple union bound, the total number of intersecting hypergraphs is at most $\binom{N}{\ell}2^{R+\ell}$. Therefore, since $N = \binom{n}{k}$,

$$\log I(n,k) \leq R + \ell + \ell \log\left(\frac{Ne}{\ell}\right) = R + \ell \log\left(\frac{2e\binom{n}{k}}{\ell}\right).$$

Because $R = (1+\varepsilon)\binom{n-1}{k-1}$, it is enough to show that $\ell \log\left(\frac{2e\binom{n}{k}}{\ell}\right) \leq \varepsilon\binom{n-1}{k-1}$. We have

$$\ell = \frac{1+\varepsilon}{\varepsilon} \cdot \frac{(n-k)\binom{n}{k}}{n\binom{n-k}{k}} \ln\left(\frac{n}{(1+\varepsilon)k}\right) \leq \frac{2\binom{n}{k}\ln n}{\varepsilon\binom{n-k}{k}} = \frac{2k\ln n}{\varepsilon n\binom{n-k}{k}} \cdot \binom{n-1}{k-1} \leq \frac{\ln n}{\varepsilon\binom{n-k}{k}} \cdot \binom{n-1}{k-1},$$

and hence

$$\ell \log\left(\frac{2e\binom{n}{k}}{\ell}\right) \leq \ell \log\left(\frac{2\varepsilon n e\binom{n-k}{k}}{(1+\varepsilon)(n-k)}\right) \leq \frac{\binom{n-1}{k-1}\ln n}{\varepsilon\binom{n-k}{k}} \cdot \log\left(4\varepsilon e\binom{n-k}{k}\right) \leq \varepsilon\binom{n-1}{k-1}, \tag{2.14}$$

provided that $\ln(n)\log\left(4\varepsilon e\binom{n-k}{k}\right) \leq \varepsilon^2\binom{n-k}{k}$. However, since $n \geq 2k+1$, we have $\binom{n-k}{k} = \binom{n-k}{n-2k} \geq \binom{n-k}{1} = n-k = \Omega(n)$, and so $\ln n \cdot \log\left(4\varepsilon e\binom{n-k}{k}\right)/\binom{n-k}{k} \to 0$ as $n \to \infty$. Thus (2.14) holds for n sufficiently large. Letting $\varepsilon \to 0$, we have $\log I(n,k) \leq (1+o(1))\binom{n-1}{k-1}$, as desired. $\square$

We conclude this section by observing that the $n \geq 2k+1$ bound in Theorem 2.6.1 is best possible. When $n = 2k$, the k-sets in $[n]$ come in $\frac{1}{2}\binom{n}{k} = \binom{n-1}{k-1}$ complementary pairs, and a hypergraph is intersecting if and only if it does not contain both edges from a single pair. We thus have $I(n,k) = 3^{\binom{n-1}{k-1}}$ when $n = 2k$. For $n < 2k$, the complete hypergraph $\binom{[n]}{k}$ is itself intersecting, and thus $I(n,k) = 2^{\binom{n}{k}}$.

Permutations

For t-intersecting families of permutations, the auxiliary graph is called the derangement graph Γ_t. The vertex set of Γ_t is S_n and two permutations $\sigma, \tau \in S_n$ are adjacent in Γ_t if and only if $|\sigma \cap \tau| < t$. Thus, t-intersecting families of permutations correspond to independent sets in Γ_t.

Let G be a group and S be a generating set. The Cayley graph $\Gamma = \Gamma(G, S)$ has a vertex set G and edge set $\left\{gh : gh^{-1} \in S\right\}$. Alternatively one can think of Γ_t as the Cayley graph on S_n generated by the set

$$FPF_t = \{\sigma \in S_n : \sigma \text{ has less than } t \text{ fixed points}\}.$$

For example, for intersecting families, Γ_1 is the Cayley graph on S_n generated by the set of fixed point free permutations, i.e. $\{\sigma \in S_n : \sigma(i) \neq i \ \forall i \in [n]\}$. We note that Γ_t is a D-regular graph with

$$D = |FPF_t| = n! \sum_{i=0}^{t-1} \frac{1}{i!} \sum_{j=0}^{n-i} \frac{(-1)^j}{j!} = \left(\frac{1}{e} + o(1)\right) n! \sum_{i=0}^{t-1} \frac{1}{i!}$$

on $N = n!$ vertices.

Ellis, Friedgut, and Pilpel [27] calculate all of the eigenvalues of Γ_t; in particular the eigenvalue of interest in applying the Expander Mixing Lemma. the second largest eigenvalue in absolute value, is the smallest eigenvalue of Γ_t:

Theorem 2.6.7. *For n sufficiently large, the least eigenvalue of Γ_t is*

$$\begin{aligned}
\lambda &= \sum_{i=0}^{t-1} \binom{n}{i} \frac{i-1}{n-1} d_{n-i} = \sum_{i=0}^{t-1} \binom{n}{i} \frac{i-1}{n-1} \sum_{j=0}^{n-i} (-1)^j \frac{n-i!}{j!} \\
&= n! \sum_{i=0}^{t-1} \frac{1}{i!} \frac{i-1}{n-1} \sum_{j=0}^{n-i} \frac{(-1)^j}{j!} = \left(\frac{1}{e} + o(1)\right) n! \sum_{i=0}^{t-1} \frac{1}{i!} \frac{i-1}{n-1} \\
&= \left(\frac{1}{e} + o(1)\right) \frac{n!}{n-1} \sum_{i=-1}^{t-2} \frac{i}{(i+1)!} = -\left(\frac{1}{e} + o(1)\right) \frac{n!}{n-1} \left(1 - \sum_{i=1}^{t-2} \frac{i}{(i+1)!}\right)
\end{aligned}$$

where d_{n-i} denotes the number of derangements of S_{n-i}.

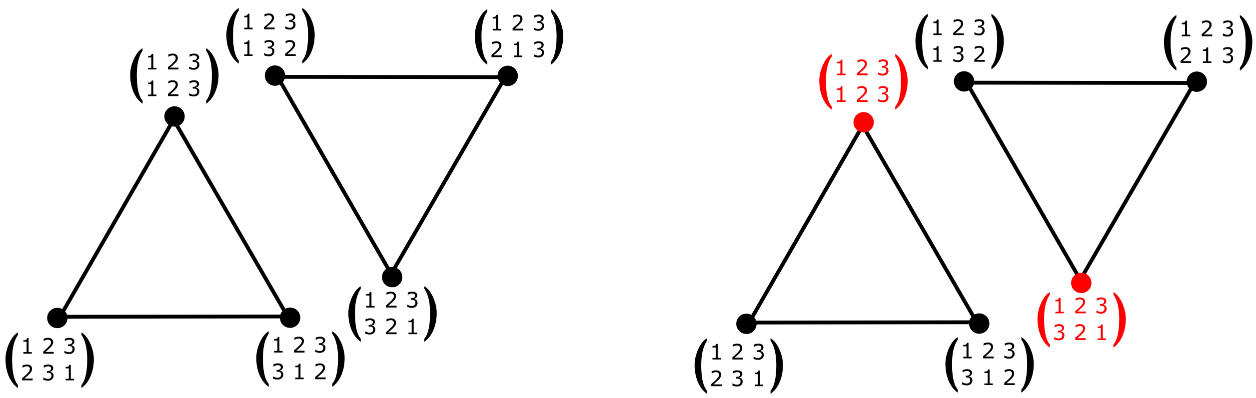

Figure 2.2: Independent Sets in Derangement Graphs

Independent sets in the derangement graph on S_n, correspond to intersecting families in S_n. Here is the derangment graph Γ_1 for $n = 3$.

The analogous main result for permutations is:

Theorem 2.6.8. *For n sufficiently large, the number of intersecting t-intersecting families of permutations in S_n is*

$$2^{(1+o(1))(n-t)!}.$$

Invoking Theorem 2.6.3, we get the following supersaturation bound.

Proposition 2.6.9. *Given $\varepsilon > 0$ and n sufficiently large, any set S of at least $(1+\varepsilon)(n-t)!$ vertices in the derangment graph Γ_t induces at least $\frac{1}{n-1}\left(\frac{1}{e}+o(1)\right)\left(\sum_{i=0}^{t-1}\frac{n-i}{i!}+\frac{n^t}{(1+\varepsilon)}\sum_{i=0}^{t-1}\frac{i-1}{i!}\right)\binom{|S|}{2}$ edges.*

Proof. Given a vertex set S with $|S| \geq (1+\varepsilon)(n-t)!$, we apply Theorem 2.6.3 and the fact that the second largest eigenvalue in absolute value is $\lambda = -\left(\frac{1}{e}+o(1)\right)\frac{n!}{n-1}\left(1-\sum_{i=1}^{t-2}\frac{i}{(i+1)!}\right)$ to obtain that

$$\begin{aligned}
e(S) &\geq \left(\frac{D-\lambda}{N}+\frac{\lambda}{|S|}\right)\binom{|S|}{2} \geq \left(\frac{1}{e}+o(1)\right)\left(\sum_{i=0}^{t-1}\frac{1}{i!}\frac{n-i}{n-1}+\frac{n!}{|S|}\sum_{i=0}^{t-1}\frac{1}{i!}\frac{i-1}{n-1}\right)\binom{|S|}{2} \\
&\geq \frac{1}{n-1}\left(\frac{1}{e}+o(1)\right)\left(\sum_{i=0}^{t-1}\frac{n-i}{i!}+\frac{n\cdot\ldots\cdot(n-t+1)}{(1+\varepsilon)}\sum_{i=0}^{t-1}\frac{i-1}{i!}\right)\binom{|S|}{2} \\
&\geq \frac{1}{n-1}\left(\frac{1}{e}+o(1)\right)\left(\sum_{i=0}^{t-1}\frac{n-i}{i!}+\frac{n^t}{(1+\varepsilon)}\sum_{i=0}^{t-1}\frac{i-1}{i!}\right)\binom{|S|}{2}.
\end{aligned}$$

□

We can combine this supersaturation result with Theorem 2.6.2 in order to find a small set of containers of independent sets in the derangement graph. From this, we will be able to derive the main result, namely Theorem 2.6.8.

Proposition 2.6.10. *For $\varepsilon > 0$ and n sufficiently large, let $R = (1+\varepsilon)(n-t)!$ and*

$$\ell = \frac{1}{\beta} \ln\left(\frac{n!}{(1+\varepsilon)(n-t)!}\right),$$

where $\beta = \frac{1}{n-1}\left(\frac{1}{e}+o(1)\right)\left(\sum_{i=0}^{t-1}\frac{n-i}{i!} + \frac{n^t}{(1+\varepsilon)}\sum_{i=0}^{t-1}\frac{i-1}{i!}\right)$. Then there exist families $\mathcal{F}_i \subset S_n$, $1 \leq i \leq \binom{n!}{\ell}$, each of size at most $R+\ell$, such that every intersecting family $\mathcal{F} \subset S_n$ is a subfamily of $\mathcal{F}_i$ for some i.

Proof. We apply Theorem 2.6.2 to the derangement graph Γ_t. By Proposition 2.6.5, condition (2.13) is satisfied by taking

$$\beta = \frac{1}{n-1}\left(\frac{1}{e}+o(1)\right)\left(\sum_{i=0}^{t-1}\frac{n-i}{i!} + \frac{n^t}{(1+\varepsilon)}\sum_{i=0}^{t-1}\frac{i-1}{i!}\right),$$

where $D = \left(\frac{1}{e}+o(1)\right)n!\sum_{i=0}^{t-1}\frac{1}{i!}$ and $N = n!$. In order to satisfy (2.12), we take

$$\ell = \frac{1}{\beta}\ln\left(\frac{N}{R}\right) = \frac{1}{\beta}\ln\left(\frac{n!}{(1+\varepsilon)(n-t)!}\right) \leq \frac{1}{\beta}\ln\left(n^t\right) \leq \frac{1}{\beta}t\ln(n).$$

Applying Theorem 2.6.2, the result follows by taking $\mathcal{F}_i$ to be the family $C_i \subset S_n$, since every t-intersecting family of permutations is an independent set of Γ_t. $\square$

2.7 Further Directions and Open Questions

Inspired by this work, in a preprint from 2017 Frankl and Kupavskii [35] showed analogous results for pairs of cross-intersecting families of hypergraphs; this work improves Theorem 2.1.4.

Two families of uniform hypergraphs $\mathcal{F} \subseteq \binom{[n]}{k}$ and $\mathcal{G} \subseteq \binom{[n]}{\ell}$ are said to be *cross-intersecting* if for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, then $F \cap G \neq \emptyset$. Note $\mathcal{F}$ is a k-uniform hypergraph on $[n]$ and $\mathcal{G}$ is an ℓ-uniform hypergraph on $[n]$ with k not necessarily equal to ℓ. Note that in this context, we do not stipulate that $\mathcal{F}$ and $\mathcal{G}$ themselves be intersecting hypergraphs; we only care about pairs with one edge from $\mathcal{F}$ and the other from $\mathcal{G}$ as well as that $\mathcal{F}$ and $\mathcal{G}$ live on the same ground set $[n]$.

We will denote the number of pairs of cross-intersecting families of hypergraphs $\mathcal{A} \subseteq \binom{[n]}{a}$ and $\mathcal{B} \subseteq \binom{[n]}{b}$ with $|\mathcal{A}| = t$ by $CI(n,a,b,t)$, and we define $CI(n,a,b) := \sum_t CI(n,a,b,t)$. Note that $CI(n,a,b,0)$ is simply $2^{\binom{n}{b}}$ because if $|\mathcal{A}| = 0$ then $\mathcal{B}$ can be any b-uniform hypergraph on $[n]$. Furthermore, we denote the number of pairs of cross-intersecting families $\mathcal{A} \subseteq \binom{[n]}{a}$ and $\mathcal{B} \subseteq \binom{[n]}{b}$ with $t_1 \leq |\mathcal{A}| \leq t_2$ by $CI(n,a,b,[t_1,t_2])$.

Using the Bollobás Set-Pairs Inequality [15], Frankl and Kupavskii [35] obtain an upper bound on the number of maximal pairs of cross-intersecting families:

Theorem 2.7.1. *Choose $a, b, n \in \mathbb{N}$ and put $c := \max\{a, b\}$, $T := \binom{n-a+b-1}{n-a}$. For*

$$n \geq a + b + 2\sqrt{c \log c} + 2 \max\{0, a - b\},$$

$a, b \to \infty$, and $b \gg \log a$ we have

$$CI(n, a, b) = (1 + \delta_{ab} + o(1))\, 2^{\binom{n}{c}}, \text{ and}$$

$$CI(n, a, b, [1, T]) = (1 + o(1)) \binom{n}{a} 2^{\binom{n}{b} - \binom{n-a}{b}},$$

where $\delta_{ab} = 1$ if $a = b$ and 0 otherwise.

Using this result, they are able to improve Theorem 2.1.4 as follows. As before, let $I(n, k)$ denote the total number of intersecting families $\mathcal{F} \subseteq \binom{[n]}{k}$. For a family $\mathcal{F} \subseteq \binom{[n]}{k}$ the *diversity* $\gamma(\mathcal{F})$ is defined as $|\mathcal{F}| - \Delta(\mathcal{F})$ where $\Delta(\mathcal{F}) := \max_{i \in [n]} |\{F : i \in F \in \mathcal{F}\}|$. Let $I(n, k, \geq t)$ denote the number of intersecting families with diversity at least t and $I(n, k, t)$ denote the number of intersecting families with diversity exactly t. Here $I(n, k, \geq 1)$ also counts the number of non-trivial intersecting families.

Theorem 2.7.2. *For $n \geq 2k + 2 + 2\sqrt{k \log k}$ and $k \to \infty$ we have*

$$I(n, k) = (n + o(1)) 2^{\binom{n-1}{k-1}}, \text{ and}$$

$$I(n, k, \geq 1) = (1 + o(1)) n \binom{n-1}{k} 2^{\binom{n-1}{k-1} - \binom{n-k-1}{k-1}}.$$

In this theorem,

$$n \binom{n-1}{k} 2^{-\binom{n-k-1}{k-1}} \leq 2^{n+1-\left(\frac{n-k}{k}\right)^{k-1}} \leq 2^{n+1-2^{n-2k}} = o(1).$$

We see that $I(n, k, \geq 1) = o\left(2^{\binom{n-1}{k-1}}\right)$ whereas $I(n, k, 0) = (n + o(1)) 2^{\binom{n-1}{k-1}}$. Do analogous results hold in other settings?

Two families of uniform hypergraphs $\mathcal{F} \subseteq \binom{[n]}{k}$ and $\mathcal{G} \subseteq \binom{[n]}{\ell}$ are said to be *cross-t-intersecting* if for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, then $|F \cap G| \geq t$. As is to be expected, when $t = 1$ the notion of cross-t-intersecting families of uniform hypergraphs coincides the cross-intersecting families of uniform hypergraphs introduced above.

Question 2.7.3. *Does a similar result hold for cross-t-intersecting families of uniform hypergraphs?*

Ellis, Friedgut, and Pilpel [27] studied a similar concept in the setting of permutations. More precisely, two families of permutations $\mathcal{F}, \mathcal{G} \subseteq S_n$ are said to be *cross-t-intersecting* if for all $F \in \mathcal{F}$ and $G \in \mathcal{G}$, then $|F \cap G| \geq t$.

Question 2.7.4. *Does the corresponding result hold for cross-t-intersecting families of permutations?*

One can pursue these ideas for various other extremal problems in discrete mathematics. For example, we say that a family of permutations of $[n]$ is *t-set-intersecting* if for every pair of permutations σ, π there is some t-set $X \subset [n]$ such that $\sigma(X) = \pi(X)$. Ellis [26] proved that for n sufficiently large, the biggest t-set-intersecting families are trivial; namely, they send a fixed set of t indices to a fixed set of t images.

Question 2.7.5. *In the setting of t-set-intersecting families of permutations, can we show that these trivial families are also typical?*

Question 2.7.6. *In what other settings (for what other discrete structures) can we show that the trivial families are also the typical ones?*

Numerous other open problems remain. One of the motivating problems behind these results was the sparse analogue of the Erdős–Ko–Rado Theorem. We show that for $3 \leq k \leq n/4$ and p not too small, the largest intersecting subhypergraphs of the random hypergraph $\mathcal{H}^k(n,p)$ are trivial with high probability. This extends previous results, which held for $k = O\left(\sqrt{n \log n}\right)$. However, there is a considerable gap between our lower bound on p in Theorem 2.4.4 and the upper bound for which it is known that the sparse Erdős–Ko–Rado theorem is false.

Question 2.7.7. *What happens in this intermediate range of probabilities?*

Different techniques will also be required to study the problem for larger k; in this direction, Hamm and Kahn [40] recently established the sparse result for $n = 2k+1$ and $p = 1-c$ for some $c > 0$.

There is also the question of obtaining the sharp asymptotics on the number of intersecting k-uniform hypergraphs, $I(n,k)$. Theorem 2.4.3 gives these asymptotics for $n \geq 3k + 8 \ln k$, showing that almost all intersecting hypergraphs are trivial. For $n \geq 2k+1$, Theorem 2.6.1 provides a slightly weaker result, showing $\log I(n,k) \approx \binom{n-1}{k-1}$. New methods will be required to obtain the asymptotics of $I(n,k)$ itself for the complete range, as our bounds on the maximal intersecting families are not strong enough to apply when $n \leq 3k$. It is worth noting that when $n = 2k+1$, the typical intersecting families are non-trivial, as the Hilton–Milner families outnumber the trivial ones. However, we suspect that $n \geq 2k+2$ may already suffice for the trivial families to become typical.

The problem of enumerating maximal intersecting structures is interesting in its own right. Here we provide reasonably sharp upper bounds through the use of the Bollobás set-pairs inequality and its variants.

We can also obtain lower bounds of a similar nature. For instance, we construct t-intersecting k-uniform hypergraphs in the following manner; for each bipartition of the form $[2(k-t)+2] = X_1 \cup X_2$, we selecting one of either $X_1 \cup [2(k-t)+3, 2k-t+1]$ or $X_2 \cup [2(k-t)+3, 2k-t+1]$ to be an edge in our hypergraph.

This procedure gives $2^{\binom{2(k-t)+1}{k-t}}$ total t-intersecting k-uniform hypergraphs, each of which can be extended to a distinct maximal t-intersecting k-uniform hypergraph over $[n]$, while Proposition 2.4.5 gives an upper bound of $2^{n\binom{2(k-t)+1}{k-t}}$. By using related constructions, we find similar lower bounds in the settings of permutations and vector spaces. We believe the lower bounds to be closer to the truth, and more refined arguments would be required in order to bridge the gap.

Chapter 3

Star Decompositions of Random Regular Graphs

3.1 Introduction

An *orientation* of a graph is an assignment of exactly one direction to each of the edges. A *flow* on a graph G is a pair (D, f) where D is an orientation of G and f is a weight function on $E(G)$ satisfying

$$\sum_{e \in \delta^+(v)} f(e) = \sum_{e \in \delta^-(v)} f(e)$$

for all $v \in V(G)$, where $\delta^+(v)$ is the set of edges out of v and $\delta^-(v)$ is the set of edges into v. For integer valued k, a *nowhere-zero k-flow* on G is a flow using weights from the set

$$\{-(k-1), -(k-2), \ldots, -2, -1, 1, 2, \ldots, k-2, k-1\}.$$

In particular, a nowhere-zero 3-flow on G is a flow using weights from the set $\{-2, -1, 1, 2\}$.

Recall that a graph is said to be *d-edge-connected* if one must remove at least d edges in order to disconnect the graph. One of the most famous open problems in structural graph theory is the Nowhere-Zero 3-Flow Conjecture by Tutte [69] from 1966:

Conjecture 3.1.1. *Every 4-edge-connected graph has a nowhere-zero 3-flow.*

A *planar graph* is a graph that can be embedded in the plane without crossings. A particular planar embedding is referred to as a *plane graph.* The *dual graph* of a plane graph G is the graph with a vertex for every face of G and an edge between two vertices when the corresponding faces are separated from each other by an edge. Tutte [68] showed that there is a close relationship between nowhere-zero flows and colorings of planar graphs. He observed that a nowhere-zero k-flow on a plane graph corresponds to a k-coloring in the dual graph and vice-versa.

In 1959 Grötzsch [38] proved:

Theorem 3.1.2. *Every triangle free (and loopless) planar graph G is 3-colorable.*

By the duality of coloring and nowhere-zero flows, this is equivalent to the following statement.

Theorem 3.1.3. *Every* 4*-edge-connected planar graph has a nowhere-zero 3-flow.*

Theorem 3.1.2 therefore shows that the Nowhere-Zero 3-Flow Conjecture is true for planar graphs.

A major breakthrough came in 2013 by L. M. Lovász, Thomassen, Wu, and Zhu [54]:

Theorem 3.1.4. *Every* 6*-edge-connected graph has a nowhere-zero 3-flow.*

However, the full conjecture still remains unsolved. As observed in [63], a graph G admits a nowhere-zero 3-flow if and only if G has an orientation in which the difference between the out-degree and in-degree of any vertex is divisible by 3. Furthermore, it is known that it is enough to prove the Nowhere-Zero 3-Flow Conjecture for 5-regular graphs; e.g., see [18] or [3]; as mentioned in [3], an equivalent statement to Conjecture 3.1.1 is:

Conjecture 3.1.5. *Every* 4*-edge-connected,* 5*-regular graph has an orientation with in-degrees 4 or 1.*

Despite some recent breakthroughs, this problem still seems intractable. In 2006 Barát and Thomassen [11] introduced a natural, related question:

Question 3.1.6. *Does every* 4*-edge-connected,* 4*-regular graph have an orientation with in-degrees 4 or 1?*

Barát and Thomassen observed that the answer in general is no (e.g., see Figure 3.1) and posed the following:

Conjecture 3.1.7. *Every* 4*-edge-connected,* 4*-regular planar graph has an orientation with in-degrees 4 or 1 (equivalently with out-degrees 0 or 3).*

An interesting observation from Bollobás's [14] book on random graphs, a typical d-regular graph is d-edge-connected; thus, translating structural problems involving d-edge-connected, d-regular graphs to the setting of random d-regular graphs is a natural idea. Thinking of results in this context provides additional machinery and helps to shed light on these problems [2]. The main result in this section, from joint work with Postle, is as follows:

Theorem 3.1.8. *A random 4-regular graph on n vertices has an orientation with out-degrees 0 or 3 asymptotically almost surely, provided that* $2n$ *is divisible by 3.*

Although at first glance this appears to be a straightforward application of the second moment method, see for instance Alon and Spencer [4]; however, this technique does not work here. Instead we use the Small Subgraph Conditioning Method of Robinson and Wormald [61]. The remainder of this chapter consists of a brief history of decomposing graphs into stars, the proof of the main result, and concludes with some open questions.

[2] Some of this work was submitted in 2016; see [21].

3.2 Orientations and Decomposing Graphs Stars

For a fixed subgraph $F \subseteq G$ an *F-decomposition* of G is a partition of the edges of G into disjoint copies of F. As Barát and Thomassen [11] note, decompositions of the edges of a graph G into copies of a small fixed subgraph can be related to orientations with certain requirements. For instance, given a 4-regular planar graph G, an orientation in G with out-degrees 0 or 3 corresponds to an S_3*-decomposition* of G.

A question that has garnered much study is whether the edges of a graph G can be decomposed into copies of a small fixed subgraph, say F. Of course, some natural divisibility conditions arise for such a decomposition, namely that $e(F)$ must divide $e(G)$. Here we are primarily interested when F is isomorphic to a star. In 1957 Kotzig [50] observed that if G is connected and $e(G)$ is even, then G decomposes into copies of S_2, the star with 2 leaves. What happens for larger stars? In particular, are there also natural conditions as to if G decomposes into copies of F when F is isomorphic to the S_3, the star with 3 leaves? Not much was known about this problem until Thomassen's [64] breakthrough results on the Weak 3-Flow Conjecture. In particular, we note the following theorem which follows from a more general theorem of L. M. Lovász, Thomassen, Wu, and Zhu [54].

Theorem 3.2.1. *If $F \simeq S_k$, the star with k leaves, and G is a d-edge-connected graph such that k divides $e(G)$ and $k \leq \lceil d/2 \rceil$, then the edge set of G decomposes into copies of F.*

We note that in fact Theorem 3.2.1 is actually tight. To see this, consider $k > \lceil d/2 \rceil$ copies of K_d with edges added so that the resulting graph G is d-regular and d-edge-connected. If there existed an S_k-*decomposition* of G, then because $k > d/2$, such a decomposition would naturally partition the vertices into $\frac{d}{2k}v(G) = \frac{d^2}{2}$ centers of the stars and $\frac{2k-d}{2k}v(G) = \frac{d(2k-d)}{2}$ non-centers. However, the non-centers must form an independent set, and thus, there are at most k of them, the desired contradiction (because $k < \frac{d(2k-d)}{2}$ when $2k - d \geq 2$).

Thus, when F is isomorphic to S_3, Theorem 3.2.1 implies that a d-regular d-edge-connected graph G has an F-decomposition if $d \geq 5$ and 3 divides $e(G)$. For $d = 3$, it is easy to observe that a 3-regular graph has an S_3-decomposition if and only if it is bipartite.

As for the case when $d = 4$, the construction in Figure 3.1 on the left provides a non-planar example of a 4-regular 4-edge-connected graph G where 3 divides $e(G)$ but G does not have an S_3-decomposition. This led Barát and Thomassen [11], who knew of this example, to conjecture in 2006 that every planar 4-regular 4-edge-connected graph G where 3 divides $e(G)$ should have an S_3-decomposition. Unfortunately in the following year, Lai [51] presented a clever counterexample (replicated in Figure 3.1 on the right) to their nice conjecture.

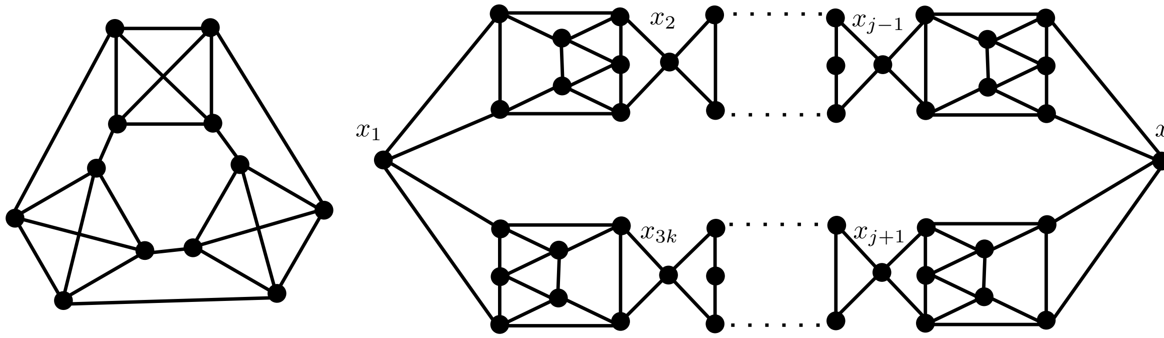

Figure 3.1: 4-Regular 4-Edge-Connected Graphs

On the left is a non-planar 4-regular 4-edge-connected graph with no S_3-decomposition. On the right is Lai's planar construction.

Given that a typical d-regular graph is d-edge-connected, a natural setting in which to study these questions is that of random regular graphs. We utilize the *configuration model* (also known as the *pairing model* in the literature) introduced by Bollobás [16] in 1980. Let $d \geq 1$ and dn be even; we take a total of dn points and partition them into n *cells* each consisting of exactly d points. Any perfect matching of $\frac{dn}{2}$ pairs of points is said to be a *configuration*, also known as a *pairing*. Each configuration corresponds to a multigraph (possibly with loops) where the cells are vertices and the pairs are edges. We denote the uniform probability space of configurations by $\mathcal{P}_{n,d}$. In the configuration model, we choose an element of $\mathcal{P}_{n,d}$ uniformly at random and discard the result if the corresponding d-regular multigraph has either loops or parallel edges. This was shown to be equivalent to choosing a d-regular (simple) graph on n vertices uniformly at random (c.f. Wormald's [74] survey paper for more details).

We consider orientations of the edges of a configuration where the out-degree of every cell is 0 or 3, where the *out-degree* of a cell is defined to be the number of points in the cell that are the tail of some edge in the orientation. We call such an orientation a $(3,0)$-*orientation*.

The main result is as follows.

Theorem 3.2.2. *A configuration in $\mathcal{P}_{n,4}$ has a $(3,0)$-orientation asymptotically almost surely, provided that $2n$ is divisible by 3.*

Any 4-regular (simple) graph G on n vertices corresponds to exactly $(4!)^n = 24^n$ configurations in $\mathcal{P}_{n,4}$. Because each such graph corresponds to the same number of configurations, it follows that G is a (uniformly) random 4-regular (simple) graph in the configuration model. The probability that a configuration in $\mathcal{P}_{n,4}$ is simple tends to a positive constant as n tends to infinity (c.f. Wormald's [74] survey paper for more details).

Thus, we have the following as a corollary.

Corollary 3.2.3. *The edges of a random 4-regular (simple) graph on n vertices can be decomposed into copies of S_3 asymptotically almost surely, provided that $2n$ is divisible by 3.*

Our proof uses the small subgraph conditioning method of Robinson and Wormald [61]. We outline the proof of our main result before proving the remaining required individual lemmas in Subsections 3.3.1, 3.3.2, and 3.3.3. However, first we note the connections between this problem and other interesting problems.

There are many connections between orientations and decompositions; of particular interest is the following conjecture of Jaeger [43] from 1984:

Conjecture 3.2.4 (Circular Flow Conjecture)**.** *Let $k \geq 3$ be odd. Every $(2k-2)$-edge-connected graph G has a* mod k-orientation*, that is, an orientation of $E(G)$ such that for every vertex the difference between its out-degree and in-degree is 0 mod k.*

In 1988 Jaeger [44] proved that his conjecture reduces to the special case of odd regular graphs as follows.

Conjecture 3.2.5 (Jaeger's Conjecture)**.** *Let $k \geq 3$ be odd. Every $(2k-2)$-edge-connected, $(2k-1)$-regular graph has a* mod k-orientation*, that is, an orientation of $E(G)$ in which every in-degree is either $(3k-1)/2$ or $(k-1)/2$.*

Note that when $k = 3$, Conjecture 3.2.4 is actually equivalent to Tutte's Nowhere-Zero 3-Flow Conjecture [69], one of the most famous open problems in structural graph theory. When $k = 5$, Jaeger's Conjecture implies the equally famous Tutte's Nowhere-Zero 5-Flow Conjecture [68]. Thomassen [64] proved Conjecture 3.2.4 for multigraphs when the edge-connectivity is at least $2k^2 + k$. L. M. Lovász, Thomassen, Wu, and Zhang [54] later improved this and proved Conjecture 3.2.4 for graphs with edge-connectivity at least $3k - 3$.

Despite these massive breakthroughs, proving Jaeger's Conjecture still seems difficult. Yet, as noted before, a typical $(2k-1)$-regular graph is $(2k-1)$-edge-connected, and therefore, a natural idea is to study Conjecture 3.2.5 instead in the setting of random $(2k-1)$-regular graphs. Using spectral techniques, Jaeger's Conjecture was confirmed to hold for random regular graphs provided that k is large enough by Alon and Prałat [3] in 2011 as follows.

Theorem 3.2.6. *For large k, Jaeger's Conjecture holds asymptotically almost surely for random $(2k-1)$-regular graphs.*

The proof makes use of the Expander Mixing Lemma and does not apply when k is too small.

Recently, utilizing the small subgraph conditioning method of Robinson and Wormald [61], Prałat and Wormald [60] were able to confirm Jaeger's Conjecture (Conjecture 3.2.5) for the case when $k = 3$, namely they proved the following theorem.

Theorem 3.2.7. *Tutte's nowhere-zero* 3*-flow conjecture holds asymptotically almost surely for random* 5*-regular graphs.*

Given these results, we were inspired to consider decompositions of random regular graphs, in particular whether the Barát–Thomassen Conjecture might hold in the random case. Given Theorem 3.2.1, it is also natural to wonder more generally whether random d-regular graphs have S_k decompositions for some $k > \lceil d/2 \rceil$. We believe our methods could be applied to these questions.

As for other subgraphs F, Barát and Thomassen [11] conjectured in 2006 that for every tree T, there exists c_T such that every c_T-edge-connected graph has a decomposition of its edges into copies of $F \simeq T$. Theorem 3.2.1 confirmed this when T is a star and indeed gives the best possible value of c_T in that case. More recently, Bensmail, Harutyunyan, Le, Merker, and Thomassé [12] proved the conjecture for all trees T. However, determining what the best possible edge-connectivity is or, in the case of random regular graphs, what the best possible degree is, are still open problems.

3.3 Proof of Main Theorem

Let $Y = Y(n)$ denote the number of $(3,0)$-orientations of a random element of $\mathcal{P}_{n,4}$. In Subsection 3.3.1, we approximate $\mathbb{E}[Y]$ using Stirling's approximation as follows.

Lemma 3.3.1.

$$\mathbb{E}[Y] \sim \frac{3}{\sqrt{2}} \left(\frac{27}{16}\right)^{n/3} > 0.$$

In order to show that configurations admit at least one $(3,0)$-orientation, we need to show that asymptotically almost surely $Y > 0$. A natural first approach is to attempt to use the second moment method (coming from Chebyshev's inequality) which says that if Y is a non-negative random variable and $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \to 0$ as $n \to \infty$, then asymptotically almost surely $Y > 0$. To that end, we approximate $\mathbb{E}[Y^2]$ in Subsection 3.3.2 using optimization, Taylor expansions, and multivariable integration to obtain the following.

Lemma 3.3.2.

$$\mathbb{E}[Y^2] \sim \frac{2\pi n}{9} \cdot \frac{81}{4\pi n}\sqrt{\frac{3}{2}} \cdot \left(\frac{27}{16}\right)^{2n/3} = \sqrt{\frac{3}{2}} \cdot \frac{9}{2}\left(\frac{27}{16}\right)^{2n/3}.$$

Unfortunately, we see that the second moment method does not apply because $E[Y]^2$ and $E[Y^2]$ are of the same order, as the following corollary notes.

Corollary 3.3.3.

$$\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \sim \sqrt{\frac{3}{2}} > 0.$$

Here we will use the small subgraph conditioning method of Robinson and Wormald [61]. We slightly reformulate the version appearing in the 1999 survey paper of Wormald [74], stated here as Theorem 3.3.4, as a “black box” for our calculations.

Theorem 3.3.4. *Let $\lambda_j > 0$ and $\delta_j > -1$ be real numbers, for all $j \geq 1$. Suppose for each n there are non-negative random variables $X_j = X_j(n)$, $j \geq 1$, and $Y = Y(n)$ (defined on the same probability space) such that X_j is integer valued and $\mathbb{E}[Y] > 0$ (for n sufficiently large). Furthermore, suppose that*

(1.) for each $j \geq 1$, $X_1, X_2, \ldots, X_j$ are asymptotically independent Poisson random variables with

$$\mathbb{E}[X_i] \to \lambda_i, \text{ for all } i \in [j];$$

(2.)

$$\frac{\mathbb{E}\left[Y[X_1]_{\ell_1} \ldots [X_j]_{\ell_j}\right]}{\mathbb{E}[Y]} \to \prod_{i=1}^{j} \left(\lambda_i \left(1 + \delta_i\right)\right)^{\ell_i}$$

for any fixed $\ell_1, \ldots, \ell_j$ where $[X]_\ell$ is the falling factorial;

(3.)

$$\sum_{i \geq 1} \lambda_i \delta_i^2 < \infty; \text{ and } \frac{\mathbb{E}[Y(n)^2]}{\mathbb{E}[Y(n)]^2} \leq \exp\left(\sum_{i \geq 1} \lambda_i \delta_i^2\right) + o(1) \text{ as } n \to \infty.$$

Then, asymptotically almost surely $Y > 0$.

Here the distribution of Y is affected by small but not too common (expected number is bounded) subgraphs of the random 4-regular graph, namely short cycles. In such situations we can attempt to apply the small subgraph conditioning method. When this method works, by conditioning on the small subgraph counts, we are able to control the variance of Y and in so doing show that $Y > 0$ asymptotically almost surely.

To understand how this works, consider partitioning the set of all 4-regular graphs on n vertices (with $2n$ divisible by 3) by the number of triangles. Within each partition class, the expected number of $(3, 0)$-orientations may be smaller than $\mathbb{E}[Y]$, though by at most a constant factor. Meanwhile the variance inside each class is smaller than the variance of Y. Applying the second moment method to the classes individually yields an increase in the probability that $Y > 0$, yet it still does not show that this probability tends to 1 asymptotically. So we further partition the classes by the number of 4-cycles, then by the number of 5-cycles, and so on. Surprisingly, by conditioning on the numbers of all cycles, it is possible to reduce the variance of Y to any desired fraction of $\mathbb{E}[Y]^2$. Intuitively, this seems plausible as graphs that have the same number of triangles, 4-cycles, etc. tend to have a similar structure and so admit less variance in the number of S_3-decompositions. Thankfully we do not actually perform such an analysis, relying on the method of Robinson and Wormald [61]; for a proof see Janson's [45] paper.

As in most applications of this method in the literature, we let X_j denote the number of cycles of length j in the multigraph corresponding to a random element of $\mathcal{P}_{n,4}$. Here, for $j \geq 1$, $X_1, X_2, \ldots, X_j$ are asymptotically independent Poisson random variables and

$$\mathbb{E}[X_j] \to \lambda_j := \frac{3^j}{2 \cdot j}.$$

This is an immediate consequence of the following theorem of Bollobás [16] from 1980:

Theorem 3.3.5. *For d fixed, let X_j denote the number of cycles of length j in the random multigraph resulting from a configuration in $\mathcal{P}_{n,d}$. For $j \geq 1$, $X_1, \ldots, X_j$ are asymptotically independent Poisson random variables with means $\lambda_j = \frac{(d-1)^j}{2 \cdot j}$.*

In Subsection 3.3.3, we compute $\mathbb{E}[YX_j]$ as follows by extending orientations of small cycles to orientations of the entire graph.

Lemma 3.3.6.

$$\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} \sim \frac{3^j}{2 \cdot j}\left(1 + \left(-\frac{1}{3}\right)^j\right) = \lambda_j \left(1 + \left(-\frac{1}{3}\right)^j\right).$$

An easy observation from the first examinations of random graphs is that, for any fixed subgraph H with more edges than vertices, a multigraph corresponding to a random element of $\mathcal{P}_{n,4}$ asymptotically almost surely contains no subgraph isomorphic to H. Informally speaking, we would not expect to have two cycles sharing edges (or for that matter vertices). Therefore, we concentrate on disjoint cycles and roughly think of them as being independent. These observations combined with Lemma 3.3.6 imply the following more general form of Lemma 3.3.6, which computes the joint factorial moments.

Corollary 3.3.7.

$$\frac{\mathbb{E}\left[Y[X_1]_{\ell_1} \ldots [X_j]_{\ell_j}\right]}{\mathbb{E}[Y]} \to \prod_{i=1}^{j}\left(\frac{3^i}{2 \cdot i}\left(1 + \left(-\frac{1}{3}\right)^i\right)\right)^{\ell_i} = \prod_{i=1}^{j}\left(\lambda_i\left(1 + \left(-\frac{1}{3}\right)^i\right)\right)^{\ell_i}$$

holds for any fixed $\ell_1, \ldots, \ell_j$ will from this.

From Lemma 3.3.6, $\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} \sim \lambda_j \left(1 + \left(-\frac{1}{3}\right)^j\right)$; thus, we set $\delta_j := -\left(\frac{1}{3}\right)^j > -1$ and verify the following.

Lemma 3.3.8.

$$\sum_{i\geq 1}\lambda_i\delta_i^2 < \infty \textit{ and } \exp\left(\sum_{i\geq 1}\lambda_i\delta_i^2\right) = \sqrt{\frac{3}{2}} \sim \frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2}.$$

Proof. Recall that $\lambda_i = \frac{3^i}{2\cdot i}$. Using that $\sum_{i\geq 1}\frac{x^i}{i} = -\ln(1-x)$ for all $-1 < x < 1$, we obtain that

$$\sum_{i\geq 1}\lambda_i\delta_i^2 = \sum_{i\geq 1}\frac{3^i}{2\cdot i}\cdot\left(-\frac{1}{3}\right)^{2i} = \frac{1}{2}\sum_{i\geq 1}\frac{\left(\frac{1}{3}\right)^i}{i} = \frac{1}{2}\left(-\ln(2/3)\right) < \infty.$$

Thus,

$$\exp\left(\sum_{i\geq 1}\lambda_i\delta_i^2\right) = \exp\left(\frac{1}{2}\left(-\ln(2/3)\right)\right) = \sqrt{\frac{3}{2}}.$$

□

Modulo proofs of Lemma 3.3.1 (proved in Subsection 3.3.1), Lemma 3.3.2 (proved in Subsection 3.3.2), and Lemma 3.3.6 (proved in Subsection 3.3.3), we see that Theorem 3.3.4 now immediately implies our main result as follows.

Proof of Main Result (Theorem 3.2.3). Let 3 divide n and $Y = Y(n)$ denote the number of $(3,0)$-orientations of a random element of $\mathcal{P}_{n,4}$. Let X_j denote the number of cycles of length j in a random element of $\mathcal{P}_{n,4}$. We now can apply Theorem 3.3.4 to Y and X_j. Note that condition (1.) holds by Theorem 3.3.5, condition (2.) holds by Corollary 3.3.7, and condition (3.) holds by Lemma 3.3.8. Thus $Y > 0$ asymptotically almost surely, as desired. □

3.3.1 Expected Number of Decompositions

We let $Y = Y(n)$ denote the number of $(3,0)$-orientations of a random element of $\mathcal{P}_{n,4}$. We will make use of the following definition. Given n cells each consisting of 4 points, a *signature* is a set of $2n/3$ points no two belonging to the same cell. We call these points the *special points* of the signature. We refer to a cell as a *center* if it contains a special point and as a *leaf* otherwise. We say a point is an *in-point* if it is special or in a leaf of the signature and say it is an *out-point* otherwise.

We say a configuration in $\mathcal{P}_{n,4}$ *extends* a signature if the configuration forms a perfect matching between the in-points and the out-points of the signature. We note that a $(3,0)$-orientation extends exactly one signature. In this signature, the *centers* correspond to the $2n/3$ cells of out-degree 3 in the orientation (here the special point in each center is the head of the only incoming edge) and the *leaves* correspond to the remaining $n/3$ cells of out-degree 0 in the orientation.

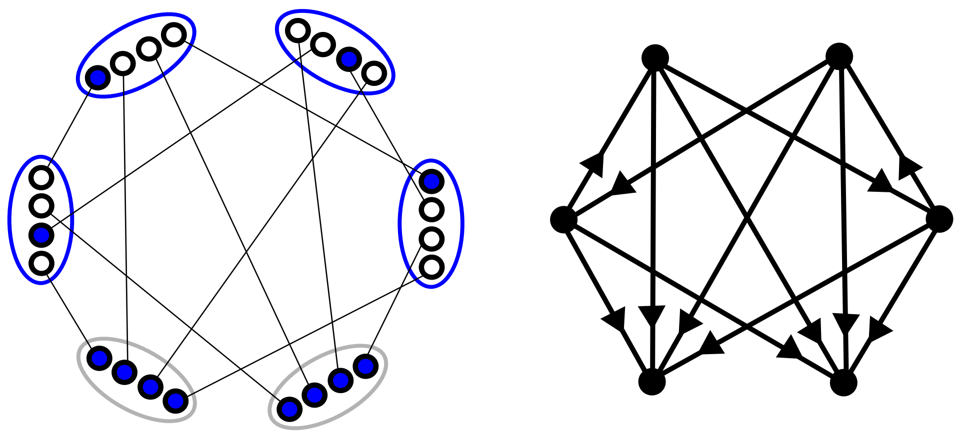

Figure 3.2: A Signature and the Corresponding $(3, 0)$-orientation
Signatures correspond to $(3, 0)$-orientations of a random element of $\mathcal{P}_{n,4}$.

To prove Lemma 3.3.1 though, we switch the order of counting and instead count the number of configurations that extend a given signature. We are now prepared to prove Lemma 3.3.1 as follows.

Proof of Lemma 3.3.1. There are a total of $\binom{n}{2n/3}4^{2n/3}$ signatures for n cells of 4 points. Recall that a configuration extends a given signature if and only if the configuration matches the in-points of the signature with its out-points. Thus, there are $(\frac{4n}{2})! = (2n)!$ configurations that extend a given signature to a $(3, 0)$-orientation. Using Stirling's approximation $s! \sim \sqrt{2\pi s}\left(\frac{s}{e}\right)^s$, we see that

$$\mathbb{E}[Y] = \frac{\binom{n}{2n/3}4^{2n/3}(2n)!}{M(4n)} = 4^{5n/3}\frac{\binom{n}{2n/3}}{\binom{4n}{2n}} \sim \frac{3}{\sqrt{2}}\left(\frac{3^3}{2^4}\right)^{n/3} = \frac{3}{\sqrt{2}}\left(\frac{27}{16}\right)^{n/3} > 0,$$

where

$$M(4n) = \frac{(4n)!}{(\frac{4n}{2})! \cdot 2^{(4n)/2}} = \frac{(4n)!}{(2n)! \cdot 2^{2n}}$$

is the number of perfect matchings of $4n$ points. □

3.3.2 The Second Moment Method

In order to calculate $\mathbb{E}[Y^2]$ for Lemma 3.3.2, we should count the number of pairs of signatures that a given configuration extends. As in the proof of Lemma 3.3.1, we invert this count by fixing a pair of signatures S_1 and S_2 and then calculating how many configurations that they both jointly extend. To

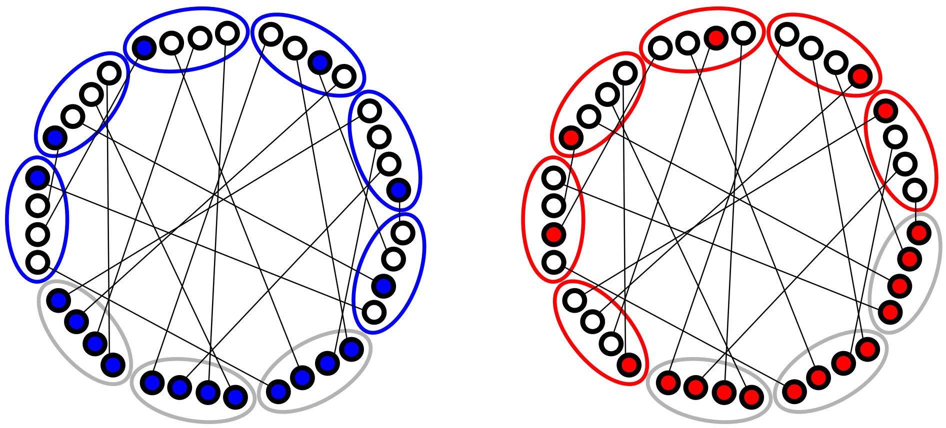

Figure 3.3: Pairs of Signatures

An example of two signatures S1 and S2 that jointly extend to the same configuration.

facilitate this count, we consider how the two signatures overlap. One might think that there would be some necessary restriction on how the signatures overlap in order to guarantee the existence of even one configuration that they jointly extend, but strangely this is not the case.

Lemma 3.3.9. *For each A and B, there are*

$$\binom{n}{A, B, \frac{2n}{3} - A - B, \frac{2n}{3} - A - B, A + B - \frac{n}{3}} \cdot 4^{2(\frac{2n}{3} - A - B)} \cdot 4^A \cdot (4 \cdot 3)^B$$

pairs of signatures such that the number of cells that are centers of both signatures with the same special point is A and the number of cells that are centers of both signatures with different special points is B.

Furthermore, for each such pair of signatures, there are

$$(3A + 2B)! \cdot (2n - 3A - 2B)!$$

configurations extending both signatures.

Proof of Lemma 3.3.9. Let A denote the number of cells that are centers in both S_1 and S_2 and have the same special point. Let B denote the number of cells that are centers in both S_1 and S_2 and have different special points. Note that $A + B$ is maximized when all centers in S_1 are centers in S_2 as well; thus, $\frac{n}{3} \leq A + B \leq \frac{2n}{3}$.

We see that we may write $\mathbb{E}[Y^2]$ in terms of n, A, and B as follows. Let C_1 denote the centers of S_1 and C_2 the centers of S_2. Note that $|C_1 \cap C_2| = A + B$. Hence $|C_1 \setminus C_2| = \frac{2n}{3} - A - B = |C_2 \setminus C_1|$. There are a total of $n - |C_1 \cup C_2| = A + B - \frac{n}{3}$ remaining cells.

Of the $4n$ total points, $3A + 2B$ are out-points of both signatures with $3A$ contained in centers with the same special point and $2B$ contained in centers with different special points. There are also $3A + 2B$ points that are in-points in both signatures with A contained in centers with the same special point in S_1 and S_2, $2\left(\frac{2n}{3} - A - B\right)$ contained in a leaf of one signature and contained in a center as a special point of the other, and $4\left(A + B - \frac{n}{3}\right)$ contained in leaves of both S_1 and S_2.

Note there are $2n - 3A - 2B$ points that are out-points in S_1 and in-points in S_2 with B contained in centers with different special points and $3\left(\frac{2n}{3} - A - B\right)$ contained in a leaf of S_2 and contained in a center but not as a special point of S_1. Similarly there are $B + 3\left(\frac{2n}{3} - A - B\right) = 2n - 3A - 2B$ points that are in-points of S_1 but out-points of S_2.

Hence, by the discussion above, we see that for each possible value of A and B, there are

$$\binom{n}{A, B, \frac{2n}{3} - A - B, \frac{2n}{3} - A - B, A + B - \frac{n}{3}}$$

ways to partition the cells into these types. There are $4^{2\left(\frac{2n}{3} - A - B\right)}$ ways to pick points that are in a leaf in one signature but a special point in the other signature. There are 4^A choices of special points from the centers in both S_1 and S_2 that have the same special point. Likewise, there are $(4 \cdot 3)^B$ ways to choose special points for the centers of both S_1 and S_2 with different special points. This proves the first assertion.

For the second assertion, note that when extending two signatures, we need to ensure that in the configuration the points that are in-points of S_1 and in-points of S_2 get matched to points that are out-points of both S_1 and S_2; there are $(3A + 2B)!$ ways to do this. Similarly, we need to ensure that the points that are in-points of S_1 and out-points of S_2 get matched to points that are out-points of S_1 but in-points of S_2 (and vice versa); there are $(2n - 3A - 2B)!$ ways to do this. $\square$

Corollary 3.3.10.

$$\mathbb{E}[Y^2] = \sum_{A,B} \frac{(2n)! \cdot n! \cdot 4^{\frac{7n}{3}} \cdot 3^B \cdot (3A + 2B)! \cdot (2n - 3A - 2B)!}{(4n)! \cdot 4^{A+B} \cdot A! \cdot B! \cdot \left(\frac{2n}{3} - A - B\right)! \cdot \left(\frac{2n}{3} - A - B\right)! \cdot \left(A + B - \frac{n}{3}\right)!}$$

where A and B are non-negative integers such that $\frac{n}{3} \leq A + B \leq \frac{2n}{3}$.

Proof of Corollary 3.3.10. The computation goes as follows. We range over all possibilities of how two signatures may overlap, i.e. we range over A and B. Using Lemma 3.3.9 and tidying the formula gives that

$$
\begin{aligned}
\mathbb{E}[Y^2] &= \frac{1}{M(4n)} \sum_{\#\text{pairs of signatures}} \#\text{configurations extending both} \\
&= \frac{2^{2n}(2n)!}{(4n)!} \sum_{A,B} \binom{n}{A, B, \frac{2n}{3} - A - B, \frac{2n}{3} - A - B, A + B - \frac{n}{3}} \\
&\cdot 4^{2\left(\frac{2n}{3} - A - B\right)} \cdot 4^A \cdot (4 \cdot 3)^B \cdot (3A + 2B)! \cdot (2n - 3A - 2B)! \\
&= \sum_{A,B} \frac{(2n)! \cdot n! \cdot 4^{\frac{7n}{3}} \cdot 3^B \cdot (3A + 2B)! \cdot (2n - 3A - 2B)!}{(4n)! \cdot 4^{A+B} \cdot A! \cdot B! \cdot \left(\frac{2n}{3} - A - B\right)! \cdot \left(\frac{2n}{3} - A - B\right)! \cdot \left(A + B - \frac{n}{3}\right)!}
\end{aligned}
$$

□

It is useful to normalize A and B by letting $a = A/n$ and $b = B/n$. We let L denote the region

$$
L := \left\{ \left(\frac{A}{n}, \frac{B}{n} \right) \in \mathbb{R}^2 : A, B \in \mathbb{Z} \cap [0, 2n/3] \text{ and } \frac{n}{3} \leq A + B \leq \frac{2n}{3} \right\}.
$$

Thus the sum in Corollary 3.3.10 ranges over L. We will also need to consider points of L but without the restriction of A and B being integral. Thus we let R denote the region

$$
R := \left\{ (a, b) \in \mathbb{R}^2 : 0 \leq a, b \leq \frac{2}{3} \text{ and } \frac{1}{3} \leq a + b \leq \frac{2}{3} \right\}.
$$

To prove Lemma 3.3.2, we will apply Stirling's formula to the formula in Corollary 3.3.10. Doing so will yield an exponential part and a polynomial part.

To that end, we introduce the two following functions. First for the exponential part, let us define

$$
\begin{aligned}
f(a, b) :=& b(\ln 3 - \ln 4) + (2 - 3a - 2b) \ln(2 - 3a - 2b) + (3a + 2b) \ln(3a + 2b) - a(\ln a + \ln 4) \\
& - b \ln b - 2 \left(\frac{2}{3} - a - b \right) \ln \left(\frac{2}{3} - a - b \right) - \left(a + b - \frac{1}{3} \right) \ln \left(a + b - \frac{1}{3} \right) - \frac{2}{3} \ln 4,
\end{aligned}
$$

and therefore,

$$
e^{f(a,b)} = \frac{2^2 \cdot 4^{\frac{7}{3}} \cdot 3^b \cdot (2 - 3a - 2b)^{(2-3a-2b)} \cdot (3a + 2b)^{(3a+2b)}}{4^4 \cdot 4^{a+b} \cdot a^a \cdot b^b \cdot \left(\frac{2}{3} - a - b\right)^{\left(\frac{2}{3} - a - b\right)} \cdot \left(\frac{2}{3} - a - b\right)^{\left(\frac{2}{3} - a - b\right)} \cdot \left(a + b - \frac{1}{3}\right)^{\left(a + b - \frac{1}{3}\right)}}.
$$

Now for the polynomial part, let us define

$$
g(a, b) := \frac{1}{2\pi} \cdot \sqrt{\frac{(3a + 2b) \cdot (2 - 3a - 2b)}{2 \cdot (a + \frac{1}{6n}) \cdot (b + \frac{1}{6n}) \cdot \left(\frac{2}{3} - a - b + \frac{1}{6n}\right)^2 \cdot \left(a + b - \frac{1}{3} + \frac{1}{6n}\right)}}.
$$

We are now ready to apply Stirling's formula to Corollary 3.3.2, where recall that the formula is $s! = \left(1 + O\left(\frac{1}{s}\right)\right)\sqrt{2\pi s}\left(\frac{s}{e}\right)^s$. In fact, we apply a variant Stirling's formula known as Gosper's formula which is $s! = \left(1 + O\left(\frac{1}{s}\right)\right)\sqrt{\pi(2s+\frac{1}{3})}\left(\frac{s}{e}\right)^s$. We do this because Stirling's formula approximates 0! as 0 instead of 1, which is unwieldy for division.

Lemma 3.3.11.

$$\mathbb{E}[Y^2] = \sum_{(a,b)\in L} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n}$$

where $S(a,b)$ is the error term arising from the applications of Gosper's formula.

Proof of Lemma 3.3.11. Using Gosper's formula $s! = \left(1 + O\left(\frac{1}{s}\right)\right)\sqrt{2\pi(s+\frac{1}{6})}\left(\frac{s}{e}\right)^s$, we have the following (where $S\left(\frac{A}{n}, \frac{B}{n}\right)$ denotes the error factor arising from using Gosper approximations in the calculation below).

Thus,

$$\begin{aligned}
\mathbb{E}[Y^2] &= \sum_{(\frac{A}{n},\frac{B}{n})\in L} \frac{(2n)!\cdot n!\cdot 4^{\frac{7n}{3}}\cdot 3^B\cdot(3A+2B)!\cdot(2n-3A-2B)!}{(4n)!\cdot 4^{A+B}\cdot A!\cdot B!\cdot\left(\frac{2n}{3}-A-B\right)!\cdot\left(\frac{2n}{3}-A-B\right)!\cdot\left(A+B-\frac{n}{3}\right)!} \\
&= \sum_{(\frac{A}{n},\frac{B}{n})\in L} S\left(\tfrac{A}{n},\tfrac{B}{n}\right)\cdot\sqrt{\frac{2^4\cdot\pi^4\cdot(2n+\frac{1}{6})\cdot(n+\frac{1}{6})\cdot(3A+2B+\frac{1}{6})\cdot(2n-3A-2B+\frac{1}{6})}{2^6\cdot\pi^6\cdot(4n+\frac{1}{6})\cdot(A+\frac{1}{6})\cdot(B+\frac{1}{6})\cdot(\frac{2n}{3}-A-B+\frac{1}{6})^2\cdot(A+B-\frac{n}{3}+\frac{1}{6})}}\cdot e^{f\left(\frac{A}{n},\frac{B}{n}\right)n} \\
&\sim \sum_{(a,b)\in L} S(a,b)\cdot\frac{1}{2\pi n}\cdot\sqrt{\frac{(3a+2b)\cdot(2-3a-2b)}{2\cdot(a+\frac{1}{6n})\cdot(b+\frac{1}{6n})\cdot\left(\frac{2}{3}-a-b+\frac{1}{6n}\right)^2\cdot\left(a+b-\frac{1}{3}+\frac{1}{6n}\right)}}\cdot e^{f(a,b)n} \\
&= \sum_{(a,b)\in L} S(a,b)\cdot\frac{g(a,b)}{n}\cdot e^{f(a,b)n}.
\end{aligned}$$

$\square$

Multivariate Calculus

In order to approximate $\mathbb{E}[Y^2]$, we first need to determine the global maximum of f on the region L and since we use continuous techniques, we will instead find the global maximum of f on R. To approximate the function f, we then will take the Taylor expansion of f around the point attaining the global maximum, since the maximum value (as we will show) is unique. We extend the definition of f continuously to the boundary to the boundary of R by defining $x\ln x := 0$ when $x = 0$. We prove the following.

Lemma 3.3.12. *The global maximum of f on the region R is $2\ln(3) - \frac{4}{3}\ln(4)$. This value is uniquely achieved at $P_0 = (a_0, b_0) = \left(\frac{1}{9}, \frac{1}{3}\right)$. Furthermore, the Hessian matrix, $D^2f(P_0)$, has determinant 81 and is negative definite.*

Proof of Lemma 3.3.12. To examine the stationary points, we start by computing the first partials of f:

$$\frac{\partial f}{\partial a} = -3\ln(2-3a-2b) + 3\ln(3a+2b) - \ln a - \ln 4 + 2\ln\left(\frac{2}{3} - a - b\right) - \ln\left(a+b-\frac{1}{3}\right)$$

$$\text{and } \frac{\partial f}{\partial b} = \ln 3 - \ln 4 - 2\ln(2-3a-2b) + 2\ln(3a+2b) - \ln b + 2\ln\left(\frac{2}{3} - a - b\right) - \ln\left(a+b-\frac{1}{3}\right).$$

By setting $\frac{\partial f}{\partial a} = 0$, exponentiating both sides, and rearranging, we obtain

$$(3a+2b)^3\left(\frac{2}{3} - a - b\right)^2 = 4a(2-3a-2b)^3\left(a+b-\frac{1}{3}\right). \tag{3.1}$$

By setting $\frac{\partial f}{\partial b} = \frac{\partial f}{\partial a}$, exponentiating both sides, and rearranging we obtain $3a(2-3a-2b) = b(3a+2b)$; this simplifies to

$$6a - 9a^2 = 9ab + 2b^2. \tag{3.2}$$

The only solutions in common to both equations (3.1) and (3.2) that lie in R are $\left(\frac{1}{9}, \frac{1}{3}\right)$ and $\left(\frac{2}{3}, 0\right)$. We next compute that $f\left(\frac{1}{9}, \frac{1}{3}\right) = 2\ln(3) - \frac{4}{3}\ln(4) \approx 0.348832$ and $f\left(\frac{2}{3}, 0\right) = \ln(3) - \frac{2}{3}\ln(4) = \frac{1}{2}\cdot f\left(\frac{1}{9}, \frac{1}{3}\right) \approx 0.174416$.

Now we examine the boundary of R. Along the segment $a = 0$ and $\frac{1}{3} \le b \le \frac{2}{3}$, the maximum value $f(0,b) \approx 0.253344$ occurs when $b \approx 0.393226$. Along the segment $a = \frac{1}{3} - b$ and $0 \le b \le \frac{1}{3}$, the maximum value $f(a,b) \approx 0.245950$ occurs when $a \approx 0.052556$ and $b \approx 0.280776$. Along the segment $b = 0$ and $\frac{1}{3} \le a \le \frac{2}{3}$, f is maximized when $a = \frac{2}{3}$; as above, $f\left(\frac{2}{3}, 0\right) \approx 0.174416$. Along the segment $a = \frac{2}{3} - b$ and $0 \le b \le \frac{2}{3}$, the maximum value of f occurs when $b = 0$; as above $f\left(\frac{2}{3}, 0\right) \approx 0.174416$.

Therefore, the unique global maximum occurs at $P_0 = (a_0, b_0) = \left(\frac{1}{9}, \frac{1}{3}\right)$. This proves the first assertion. Note that this corresponds to setting $3a + 2b = 2 - 3a - 2b$ and therefore $3A + 2B = 2n - 3A - 2B = n$. In other words, the number of points that are in-points of one signature and out-points of the other is equal to the number of points that are either in-points of both signatures or out-points of both signatures.

To compute the Hessian, first we take second partials of f and evaluate at P_0:

$$\frac{\partial^2 f}{\partial a^2} = \frac{9}{2-3a-2b} + \frac{9}{3a+2b} - \frac{1}{a} - \frac{2}{\frac{2}{3} - a - b} - \frac{1}{a+b-\frac{1}{3}},$$

$$\frac{\partial^2 f}{\partial a \partial b} = \frac{\partial^2 f}{\partial b \partial a} = \frac{6}{2-3a-2b} + \frac{6}{3a+2b} - \frac{2}{\frac{2}{3} - a - b} - \frac{1}{a+b-\frac{1}{3}},$$

$$\text{and } \frac{\partial^2 f}{\partial b^2} = \frac{4}{2-3a-2b} + \frac{4}{3a+2b} - \frac{1}{b} - \frac{2}{\frac{2}{3} - a - b} - \frac{1}{a+b-\frac{1}{3}}.$$

Note that

$$\frac{\partial f}{\partial a}\Big|_{P_0} = 0, \frac{\partial f}{\partial b}\Big|_{P_0} = 0, \frac{\partial^2 f}{\partial a^2}\Big|_{P_0} = -9, \frac{\partial^2 f}{\partial a \partial b}\Big|_{P_0} = \frac{\partial^2 f}{\partial b \partial a}\Big|_{P_0} = -6, \text{ and } \frac{\partial^2 f}{\partial b^2}\Big|_{P_0} = -13.$$

Thus, Hessian matrix of f evaluated at P_0 is

$$H := D^2 f(P_0) = \begin{pmatrix} -9 & -6 \\ -6 & -13 \end{pmatrix}.$$

The determinant of H is 81 with eigenvalues $-11+2\sqrt{10}$ and $-11-2\sqrt{10}$; thus, H is negative definite (this also implies that P_0 must be a local maximum). $\square$

Integrating

We are ready to prove Lemma 3.3.2. We use a Taylor expansion around P_0 combined with multivariable Gaussian integrals to calculate $\mathbb{E}[Y^2]$ more precisely.

Proof of Lemma 3.3.2. As before, we denote the Hessian matrix of f evaluated at P_0 by

$$H := D^2 f(P_0) = \begin{pmatrix} -9 & -6 \\ -6 & -13 \end{pmatrix}.$$

We denote the gradient vector of f evaluated at P_0 by

$$D := Df(P_0) = \left[\frac{\partial f}{\partial a}\Big|_{P_0}, \frac{\partial f}{\partial b}\Big|_{P_0}\right] = [0, 0].$$

We integrate near the maximum using a second-order Taylor series expansion. Let $[P - P_0]$ denote a row vector with components $[(a-a_0), (b-b_0)] = [(a-\frac{1}{9}), (b-\frac{1}{3})]$, and let $[P-P_0]^T$ be the transpose, a column vector. This gives that $f(a,b) = f(P)$ near P_0 is

$$\begin{aligned} f(P) &= f(P_0) + D[P-P_0]^T + \frac{1}{2}[P-P_0]H[P-P_0]^T + O\left(\|P-P_0\|^3\right) \\ &= f(P_0) + \frac{1}{2}[P-P_0]H[P-P_0]^T + O\left(\|P-P_0\|^3\right). \end{aligned}$$

By Taylor's Theorem, we note that the error is valid provided that $\|P - P_0\| = o(1)$. As below, we see that

$$\mathbb{E}[Y^2] \sim \frac{2\pi n}{9} \cdot \frac{g(P_0)}{n} \cdot e^{f(P_0)n} = \frac{2\pi n}{9} \cdot \frac{81}{4\pi n} \cdot \sqrt{\frac{3}{2}} \cdot \left(\frac{27}{16}\right)^{2n/3} = \sqrt{\frac{3}{2}} \cdot \frac{9}{2}\left(\frac{27}{16}\right)^{2n/3},$$

and therefore, $\frac{\mathbb{E}[Y^2]}{\mathbb{E}[Y]^2} \sim \sqrt{\frac{3}{2}}$.

By Lemma 3.3.11,

$$\mathbb{E}[Y^2] = \sum_{(a,b)\in L} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n},$$

where $S(a,b)$ is the error term arising from the applications of Gosper's formula.

Let $R' = \left\{P \in R : \|P - P_0\| = o(n^{-1})\right\}$ and $L' = L \cap R'$. Note that for all $P \in R'$, $S(P)g(P) \sim g(P_0)$ because $S(P) \sim 1$ and $\|P - P_0\|^3 = o(n^{-3})$. Here

$$g(P_0) = \frac{1}{2\pi} \cdot \sqrt{\frac{1 \cdot 1}{2 \cdot \frac{1}{3} \cdot \frac{1}{9} \cdot \left(\frac{2}{9}\right)^2 \cdot \frac{1}{9}}} = \frac{1}{2\pi} \cdot \sqrt{\frac{3 \cdot 9^4}{8}} = \frac{81}{4\pi} \cdot \sqrt{\frac{3}{2}}.$$

Thus,

$$\sum_{P\in L'} S(P) \cdot \frac{g(P)}{n} \cdot e^{f(P)n} \sim \frac{g(P_0)}{n} \cdot \sum_{P\in L'} e^{f(P)n} \sim \frac{g(P_0)}{n} \cdot e^{f(P_0)n} \cdot \sum_{P\in L'} \exp\left(\frac{1}{2}[P-P_0]H[P-P_0]^T n\right)$$
$$= \left(\frac{27}{16}\right)^{2n/3} \cdot \frac{81}{4\pi n} \cdot \sqrt{\frac{3}{2}} \cdot \sum_{P\in L'} \exp\left(\frac{1}{2}[P-P_0]H[P-P_0]^T n\right).$$

We note then that if we divide the sum by a factor of n^2, then this becomes a Riemann sum over R'. That Riemann sum in turn approximates an integral as $n \to \infty$. Hence

$$\sum_{P\in L'} \exp\left(\frac{1}{2}[P-P_0]H[P-P_0]^T n\right) = n^2 \cdot \int\int_{P\in R'} \exp\left(\frac{1}{2}[P-P_0]H[P-P_0]^T n\right) dP.$$

Now we change variables by letting $x = (a - \frac{1}{9})\sqrt{n}$ and $y = (b - \frac{1}{3})\sqrt{n}$ for $P = (a,b)$. Note that the region of x, y corresponding to R' is the whole real plane. Thus, this change of variable transforms the integral into

$$\int\int_{P\in R'} \exp\left(\frac{1}{2}[P-P_0]H[P-P_0]^T n\right) dP \sim \frac{1}{n} \cdot \int\int_{\mathbb{R}^2} \exp\left([x,\ y]\frac{H}{2}[x,\ y]^T\right) dxdy,$$

where H is the Hessian matrix of f evaluated at P_0. Diagonalizing and using the Gaussian integral, that is $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$, we see that the integral evaluates to

$$\frac{1}{n} \cdot \sqrt{\frac{\pi^2}{|\det \frac{H}{2}|}} = \frac{1}{n} \cdot \sqrt{\frac{4 \cdot \pi^2}{|\det H|}} = \frac{2\pi}{9n},$$

since $\det H = 81$ by Lemma 3.3.12. Therefore,

$$\sum_{P \in L'} S(P) \cdot \frac{g(P)}{n} \cdot e^{f(P)n} \sim \frac{2\pi}{9n} \cdot \frac{81n}{4\pi} \cdot \sqrt{\frac{3}{2}} \cdot \left(\frac{27}{16}\right)^{2n/3} = \sqrt{\frac{3}{2}} \cdot \frac{9}{2} \left(\frac{27}{16}\right)^{2n/3}.$$

Because H is negative definite, the value of f on the boundary of R' is $f(P_0) - \Omega(n^{-2})$. However, f is independent of n and P_0 is a global maximum. Thus,

$$\max_{P \in R \setminus R'} f(P) = f(P_0) - \Omega(n^{-2}) = 2\ln(3) - \frac{4}{3}\ln(4) - \Omega(n^{-2}).$$

Observe that

$$\begin{aligned}
\mathbb{E}[Y^2] &= \sum_{(a,b) \in L} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} \\
&= \sum_{(a,b) \in L \setminus L'} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} + \sum_{(a,b) \in L'} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} \\
&\sim \sum_{(a,b) \in R \setminus R'} S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} + \sqrt{\frac{3}{2}} \cdot \frac{9}{2} \left(\frac{27}{16}\right)^{2n/3}.
\end{aligned}$$

Now consider $P = (a,b) \in L \setminus L'$. Since $L \setminus L' \subseteq R \setminus R'$, we have that $e^{f(a,b)n} = \left(\frac{27}{16}\right)^{2n/3} \cdot \exp\left(-\Omega\left(n^{1/2}\right)\right)$. Yet $S(a,b) = O(1)$ and $g(a,b) = O(n^{5/2})$ as each of the terms in the denominator of g are $O(n)$.

Thus for each $(a,b) \in L \setminus L'$, we see that

$$S(a,b) \cdot \frac{g(a,b)}{n} \cdot e^{f(a,b)n} = \left(\frac{27}{16}\right)^{2n/3} \cdot \exp\left(-\Omega\left(n^{1/2}\right)\right).$$

Note that as there are only a polynomial number of points in $L \setminus L'$, namely at most n^2, the sum over points in $L \setminus L'$ is also $\left(\frac{27}{16}\right)^{2n/3} \cdot \exp\left(-\Omega\left(n^{1/2}\right)\right)$. Therefore, $\mathbb{E}[Y^2] \sim \sqrt{\frac{3}{2}} \cdot \frac{9}{2} \left(\frac{27}{16}\right)^{2n/3}$, as desired.

□

3.3.3 Joint Factorial Moments

In this section, we prove Lemma 3.3.6. By definition,

$$\mathbb{E}[YX_j] = \frac{1}{M(4n)} \sum_{j\text{-cycle } C} (\#\text{ orientations of cycle } C) \cdot (\#\text{ extensions of orientations of } C).$$

Note that this is equivalent to

$$\mathbb{E}[YX_j] = \frac{1}{M(4n)} \sum_{\text{oriented } j\text{-cycle } C} \# \text{ extensions of orientations of } C.$$

By counting how many configurations extends such oriented cycles, we will prove the following.

Lemma 3.3.13. *The number of oriented cycles with s sinks and sources is*

$$\frac{[n]_j}{j}\binom{j}{2s}(4\cdot 3)^j,$$

while the number of extensions to $(3,0)$-orientations for any oriented cycle is

$$\binom{n-j}{\frac{2n}{3}-j+s}4^{\frac{2n}{3}-j+s}\cdot 2^s\,(2n-j)!$$

Therefore, we have that

$$\mathbb{E}[YX_j] = \frac{[n]_j}{M(4n)\cdot j}4^{2n/3}3^j\sum_{s=0}^{\lfloor j/2\rfloor}\binom{j}{2s}\binom{n-j}{\frac{2n}{3}-j+s}2^{3s}(2n-j)!$$

Proof. Any oriented cycle must have the same number of sources (vertices with out-degree equal to 2) and sinks (vertices with out-degree equal to 0); an oriented cycle of length j can have s sources, s sinks and $j-2s$ other vertices for some $0 \le s \le \lfloor \frac{j}{2} \rfloor$. The number of oriented cycles of length j with exactly s sources and sinks is

$$\frac{[n]_j}{j}\binom{j}{2s}(4\cdot 3)^j.$$

To see this, choose a set of j vertices ($\binom{n}{j}$ ways). The number of cyclic permutations of j entries is $\frac{(j-1)!}{2}$, where we divide by 2 for reversing the cycle and we can pick sources and sinks in $2\binom{j}{2s}$ ways (the sources and sinks must alternate around the cycle). Finally, every vertex in the cycle needs to pick two of its 4 points in ordered fashion from the configuration for endpoints of edges in the cycle.

Let C be a cycle of length j; vertices that are sinks in C cannot have out-degree 3 and so are not centers. All other vertices must be centers. Thus the number of leaves in C is s and the number of centers in C is $j-s$. The number of extensions is then given by first completing the signature.

To this end, we choose $\frac{2n}{3}-(j-s)$ of the $n-j$ vertices outside of C to be centers and choose a special point for each such center. For each source, for the non-cycle edges we must orient one edge out and one edge in which gives 2^s choices. The cycle is already matched so this gives that the number of extensions to

$(3,0)$-orientations for any oriented cycle is

$$\binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s \left(\frac{4n-2j}{2}\right)! = \binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s\,(2n-j)!$$

Therefore the whole expression is

$$\begin{aligned}
\mathbb{E}[YX_j] &= \frac{1}{M(4n)} \sum_{s=0}^{\lfloor j/2 \rfloor} \frac{[n]_j}{j} \binom{j}{2s} (4\cdot 3)^j \binom{n-j}{\frac{2n}{3}-j+s} 4^{\frac{2n}{3}-j+s} \cdot 2^s (2n-j)! \\
&= \frac{[n]_j}{M(4n)\cdot j} 4^{2n/3} 3^j \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \binom{n-j}{\frac{2n}{3}-j+s} 2^{3s} (2n-j)!
\end{aligned}$$

$\square$

Recall that

$$\mathbb{E}[Y] = \frac{\binom{n}{2n/3} 4^{2n/3} (2n)!}{M(4n)}.$$

We are now ready to prove Lemma 3.3.6 as follows.

Proof of Lemma 3.3.6. In the computation of $\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]}$, we use the following approximation where y is a constant and x goes to infinity,

$$\frac{x!}{(x-y)!} \sim \frac{\sqrt{2\pi x}\left(\frac{x}{e}\right)^x}{\sqrt{2\pi (x-y)}\left(\frac{x-y}{e}\right)^{x-y}} \sim \left(\frac{x}{e}\right)^y \cdot \left(\frac{x}{x-y}\right)^{x-y} = \left(\frac{x}{e}\right)^y \cdot \left(1+\frac{y}{x-y}\right)^{x-y} \sim \left(\frac{x}{e}\right)^y \cdot e^y = x^y.$$

Thus,

$$\begin{aligned}
\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} &= \frac{n!}{j\cdot (n-j)!} 3^j \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \frac{\binom{n-j}{\frac{2n}{3}-j+s}}{\binom{n}{2n/3}} \frac{(2n-j)!}{(2n)!} 2^{3s} \\
&= \frac{3^j}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \frac{\left(\frac{2n}{3}\right)!}{\left(\frac{2n}{3}-j+s\right)!} \frac{\left(\frac{n}{3}\right)!}{\left(\frac{n}{3}-s\right)!} \frac{(2n-j)!}{(2n)!} 2^{3s} \\
&\sim \frac{3^j}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} \left(\frac{2n}{3}\right)^{j-s} \left(\frac{n}{3}\right)^s \frac{2^{3s}}{(2n)^j} = \frac{1}{j} \sum_{s=0}^{\lfloor j/2 \rfloor} \binom{j}{2s} 2^{2s}.
\end{aligned}$$

Note that $\binom{j}{2s}$ is the coefficient of x^{2s} in $q(x) := (1+x)^j$, so

$$\frac{\mathbb{E}[YX_j]}{\mathbb{E}[Y]} \sim \frac{1}{j}\sum_{s=0}^{\lfloor j/2\rfloor}\binom{j}{2s}2^{2s} = \frac{1}{j}\cdot\frac{(q(2)+q(-2))}{2} = \frac{1}{2\cdot j}\left(3^j+(-1)^j\right)$$
$$= \frac{3^j}{2\cdot j}\left(1+\left(-\frac{1}{3}\right)^j\right) = \lambda_j\left(1+\left(-\frac{1}{3}\right)^j\right).$$

□

3.4 Further Directions and Open Questions

Our main result implies that a random 4-regular graph has an S_3-decomposition asymptotically almost surely. A natural question is: do random d-regular graphs (with the correct edge divisibility) have edge decompositions into stars S_k?

Conjecture 3.4.1. *A random d-regular graph has a S_k-decomposition asymptotically almost surely for $k < \frac{d}{2} + O(\log d)$, provided that dn is even and $\frac{dn}{2}$ is divisible by k.*

In collaboration with Postle and Lidický, we were able to find partial evidence supporting our conjecture for small values of d using a computer:

Theorem 3.4.2. *For $d \leq 50$, a random d-regular graph has a S_k-decomposition asymptotically almost surely for $k < \frac{d}{2} + 2\log d - c$, provided that dn is even and $\frac{dn}{2}$ is divisible by k.*

For $k \leq \lceil \frac{d}{2} \rceil$, that d-edge connected graphs (with the correct edge divisibility) have S_k-decompositions follows from a result of L. M. Lovász, Thomassen, Wu, and Zhu [54].

One could also study Jaeger's Conjecture [44] from 1988 in the random setting; recall that:

Conjecture 3.4.3. *Let $k \geq 3$ be odd. Every $(2k-2)$-edge-connected $(2k-1)$-regular graph has an orientation in which every out-degree is either $\frac{3k-1}{2}$ or $\frac{k-1}{2}$.*

Using spectral methods, Alon and Prałat showed for large k Jaeger's Conjecture holds asymptotically almost surely for random $(4k+1)$-regular graphs [3]. Prałat and Wormald showed using the small subgraph conditioning method that Tutte's 3-flow conjecture (Jaeger's Conjecture for $k = 3$) holds asymptotically almost surely for random 5-regular graphs [60]. For values of k in between, this question remains open.

Chapter 4

Highly Connected, Spanning, Bipartite Subgraphs

4.1 Introduction

In a survey paper from 1989, Thomassen [65] compiled a list of conjectures fundamental to structural graph theory. Many of the problems listed are very natural and important lines of investigation. This chapter focuses in particular on Conjecture 7 from this survey.

As mentioned in [66], Erdős made the following observation in 1976:

Theorem 4.1.1. *Any graph G with minimum degree $2k-1$ contains a spanning, bipartite subgraph $H \subseteq G$ such that the minimum degree of H is at least k.*

Proof. The proof for this fact is obtained by taking a maximal edge-cut, a partition of $V(G)$ into two sets A and B, such that the number of edges with one endpoint in A and one in B, denoted $|E(A,B)|$, is maximal. Observe that if some vertex v in A does not have degree at least k in $G[B]$, then by moving v to B, one would increase $|E(A,B)|$, contrary to maximality. The same argument holds for vertices in B. In fact this proves that for each vertex $v \in V(G)$, by taking such a subgraph H, the degree of v in H, denoted $d_H(v)$, is at least $d_G(v)/2$. □

This same simple idea will be utilized throughout this chapter.

Recall that a graph is said to be *k-edge-connected* if one must remove at least k edges in order to disconnect the graph. Thomassen [66] observed that the same proof shows the following stronger statement:

Theorem 4.1.2. *If G is a graph which is at least $(2k-1)$-edge-connected, then G contains a spanning, bipartite subgraph $H \subseteq G$ such that k-edge-connected.*

In fact, each edge-cut keeps at least half of its edges.

This observation led Thomassen to conjecture that a similar phenomenon also holds for vertex-connectivity. Before proceeding to the statement of Thomassen's Conjecture, we remind the reader that a graph G is said to be *k vertex-connected* or *k-connected* if one must remove at least k vertices from $V(G)$ in order to disconnect the graph (or to remain with one single vertex). We also let $\kappa(G)$ denote the minimum integer k

for which G is k-connected. Roughly speaking, Thomassen conjectured that any graph with high enough connectivity also should contain a k-connected spanning, bipartite subgraph. The following appears as Conjecture 7 in the survey paper by Thomassen [65]:

Conjecture 4.1.3. *For all k, there exists a function $f(k)$ such that for all graphs G, if the vertex-connectivity $\kappa(G) \geq f(k)$, then there exists a spanning, bipartite subgraph $H \subseteq G$ such that $\kappa(H) \geq k$.*

In joint work with Ferber[3], we prove that Conjecture 7 holds up to a $\log n$ factor; the key idea is to carefully construct an auxiliary digraph to show the following:

Theorem 4.3.1. *For all k and n, and for every graph G on n vertices the following holds. If $\kappa(G) \geq 10^{10}k^3 \log n$, then there exists a spanning, bipartite subgraph $H \subseteq G$ such that $\kappa(H) \geq k$.*

4.2 Preliminary Results

This section introduces a number of preliminary results which will utilize in the proof of the main result in the next section.

4.2.1 Mader's Theorem

A useful tool for finding highly connected (but necessarily spanning or bipartite) subgraphs was developed by Mader [57] in 1972.

Theorem 4.2.1. *Every graph of average degree at least 4ℓ has an ℓ-connected subgraph.*

Because we are interested specifically in finding bipartite subgraphs with high connectivity, the following corollary will be helpful.

Corollary 4.2.2. *Every graph G with average degree at least 8ℓ contains a (not necessarily spanning) bipartite subgraph H which is at least ℓ-connected.*

Proof. Let G be such a graph and let $V(G) = A \cup B$ be a partition of $V(G)$ such that $|E(A, B)|$ is maximal. Observe that $|E(A, B)| \geq |E(G)|/2$, and therefore, the bipartite graph G' with parts A and B has average degree at least 4ℓ. Now, by applying Theorem 4.2.1 to the graph G' we are able to obtain the desired subgraph H. □

[3]Some of this work appeared in the Electronic Journal of Combinatorics in 2015; see [20].

4.2.2 Merging k-connected Graphs

We can use highly connected graphs as building blocks to create larger graphs which are also highly connected. We will use the following easy expansion lemma.

Lemma 4.2.3. *Let H_1 and H_2 be two vertex-disjoint graphs, each of which is k-connected. Let H be a graph obtained by adding k independent edges between these two graphs. Then, $\kappa(H) \geq k$.*

Proof. Note first that by construction, one cannot remove all the edges between H_1 and H_2 by deleting fewer than k vertices. Moreover, because H_1 and H_2 are both k-connected, each will remain connected after deleting less than k vertices. From here, the rest of the proof follows easily. □

Next we will show how to merge a collection of a few k-connected components and isolated vertices into a larger k-connected component. Before stating the next lemma formally, we will need to introduce some notation. Let $G_1, \dots, G_t$ be a collection of t vertex-disjoint k-connected graphs, let $U = \{u_{t+1}, \dots, u_{t+s}\}$ be a set consisting of a total of s isolated vertices which are disjoint to $V(G_i)$ for $1 \leq i \leq t$, and let R be a k-connected graph on the vertex set $[t+s] := \{1, \dots, t+s\}$. Finally, we let $X = (G_1, \dots G_t, u_{t+1}, \dots, u_{t+s})$ be a $(t+s)$-tuple where we use X_i to denote the ith element of tuple X. Finally, let $\mathcal{F}_R := \mathcal{F}_R(X)$ denote the family consisting of all graphs G which satisfy the following properties:

(i) the disjoint union of the elements of X is a spanning subgraph of G, and

(ii) for every distinct $i, j \in V(R)$ if $ij \in E(R)$, then there exists an edge in G between X_i and X_j, and

(iii) for every $1 \leq i \leq t$, there is a set of k independent edges between $V(G_i)$ and k distinct vertex sets $\{V(X_{j_1}), \dots, V(X_{j_k})\}$, where $V(u_i) = \{u_i\}$.

Lemma 4.2.4. *Let $G_1, \dots, G_t$ be t vertex-disjoint graphs, each of which is k-connected, and let $U = \{u_{t+1}, \dots, u_{t+s}\}$ be a set of s vertices for which $U \cap V(G_i) = \emptyset$ for every $1 \leq i \leq t$. Let R be a k-connected graph on the vertex-set $\{1, \dots, t+s\}$, and let $X = \{G_1, \dots G_t, u_{t+1}, \dots, u_{t+s}\}$. Then, any graph $G \in \mathcal{F}_R(X)$ is k-connected.*

Proof. Let $G \in \mathcal{F}_R(X)$, and let $S \subseteq V(G)$ be a subset of size at most $k-1$. We wish to show that the graph $G' := G \setminus S$ is still connected. Let $x, y \in V(G')$ be two distinct vertices in G'; we show that there exists a path in G' connecting x to y. Towards this end, we first note that if both x and y are in the same G_i, then because each G_i is k-connected, there is nothing to prove. Moreover, if both x and y are in distinct elements of X which are also disjoint from S, then we are also finished, as follows. Because R is k-connected, if we

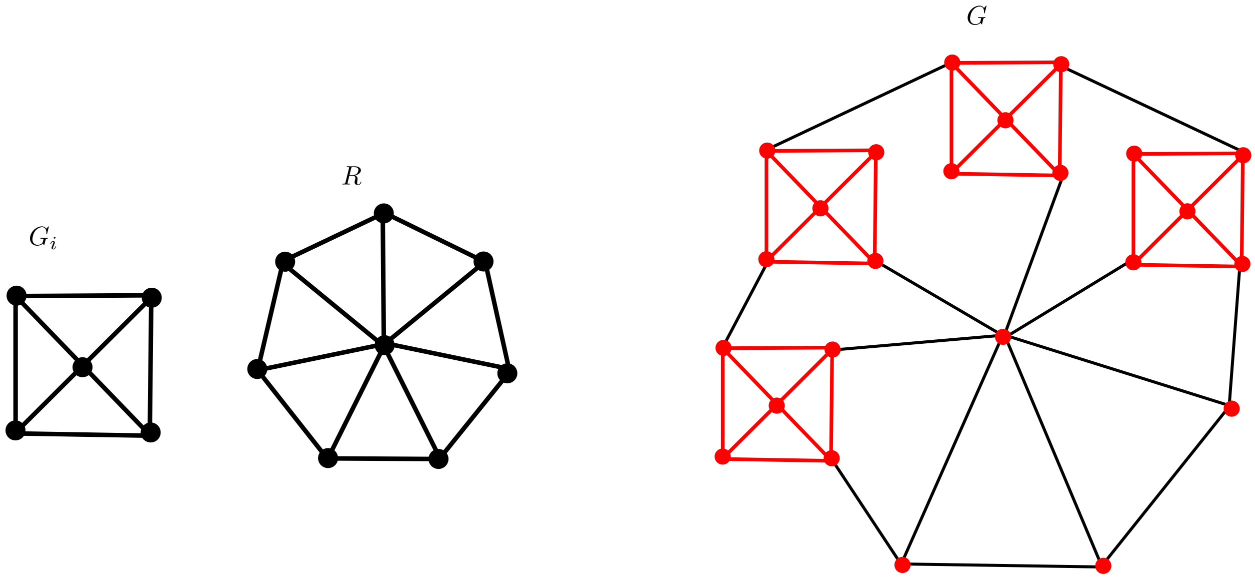

Figure 4.1: Graphs in $\mathcal{F}_R(X)$

Here we let $k = 3$ and $s = t = 4$. If R is as above, u_5, u_6, u_7, and u_8 are isolated vertices, and $G_i \simeq G_1 \simeq G_2 \simeq G_3 \simeq G_4$, then $G \in \mathcal{F}_R(X)$, where $X = (G_1, G_2, G_3, G_4, u_5, u_6, u_7, u_8)$.

delete all of the vertices in R corresponding to elements of X which intersect S, the resulting graph is still connected. Therefore, one can easily find a path between the elements containing x and y which goes only through "untouched" elements of X, and hence, there exists a path connecting x and y.

The remaining case to deal with is when x and y are in different elements of X, and at least one of them is not disjoint with S. Assume x is in some such X_i (y will be treated similarly). Using Property (iii) of $\mathcal{F}_R$, there is at least one edge between X_i and an untouched X_j. Therefore one can find a path between x and some vertex x' in an untouched X_j. This takes us back to the previous case. □

4.3 Proof of Main Result

A *directed graph* or *digraph* is a set of vertices and a collection of directed edges; note that bidirectional edges are allowed. For a directed graph D and a vertex $v \in V(D)$ we let $d_D^+(v)$ denote the out-degree of v. We let $U(D)$ denote the *underlying graph of* D, that is the graph obtained by ignoring the directions in D and merging multiple edges.

The main result of this chapter is the following theorem:

Theorem 4.3.1. *For all k and n, and for every graph G on n vertices the following holds. If $\kappa(G) \geq 10^{10}k^3 \log n$, then there exists a spanning, bipartite subgraph $H \subseteq G$ such that $\kappa(H) \geq k$.*

In order to find the desired spanning, bipartite k-connected subgraph in Theorem 4.3.1, we look at sub-digraphs in an auxiliary digraph.

The following is our main technical lemma and is the main reason why we fundamentally are left with a $\log n$ factor in the main theorem:

Lemma 4.3.2. *If D is a finite digraph on at most n vertices with minimum out-degree*

$$\delta^+(D) > (k-1)\lceil \log n \rceil,$$

then there exists a sub-digraph $D' \subseteq D$ such that

1. *For all $v \in V(D')$ we have $d^+_{D'}(v) \geq d^+_D(v) - (k-1)\lceil \log n \rceil$, and*

2. *$\kappa(U(D')) \geq k$.*

Proof. If $\kappa(U(D)) \geq k$, then there clearly is nothing to prove. So we may assume that $\kappa(U(D)) \leq k-1$. Delete a separating set of size at most $k-1$. The smallest component, say C_1, has size at most $n/2$ and for any $v \in V(C_1)$, every out-neighbor of v is either in $V(C_1)$ or in the separating set that we removed, and so

$$d^+_{C_1}(v) \geq d^+_D(v) - (k-1).$$

We continue by repeatedly applying this step, and note that this process must terminate. Otherwise, after at most $\log n$ steps we are left with a component which consists of one single vertex and yet contains at least one edge, a contradiction. $\square$

With the preliminaries out of the way, we are now ready to prove the main result, namely Theorem 4.3.1.

Proof. Let G be a finite graph on n vertices with

$$\kappa(G) \geq 10^{10} k^3 \log n.$$

In order to find the desired subgraph, we first initiate $G_1 := G$ and start the following process, finding a sequence of disjoint highly connected large bipartite subgraphs.

As long as G_i contains a bipartite subgraph which is at least k-connected on at least $10^3 k^2 \log n$ vertices, let $H_i = (S_i \cup T_i, E_i)$ be such a subgraph of maximum size, and let $G_{i+1} := G_i \setminus V(H_i)$. Note that H_1 must exist as

$$\delta(G_1) \geq 10^{10} k^3 \log n - 2k \geq 8000 k^2 \log n,$$

and so by Corollary 4.2.2, we see that G_1 must contain a k-connected bipartite subgraph of size at least $10^3k^2 \log n$.

Let $H_1, \ldots, H_t$ be the sequence obtained in this manner, and note that all the H_i's are vertex disjoint with the properties that $\kappa(H_i) \geq k$ and $|V(H_i)| \geq 10^3k^2 \log n$. In the event that at the first step H_1 is spanning, then we are finished because we have found our desired spanning highly connected bipartite subgraph of the graph G. Therefore, we proceed by supposing for a contradiction that H_1 is not actually a spanning subgraph.

Let $V_0 := V(G_{t+1}) = \{v_1, \ldots, v_s\}$ be the subset of $V(G)$ remaining after running this entire process; note that it might be the case that $V_0 = \emptyset$. Because each H_i is a bipartite, k-connected subgraph of G_i of maximum size and G is $10^{10}k^3 \log n$ connected, we show that the following are true:

(a) For every $1 \leq i < j \leq t$, there are at least $4k$ independent edges between H_i and H_j, and

(b) for every $j > i$ and $v \in V(G_j)$, the number of edges in G between v and H_i, denoted by $d_G(v, V(H_i))$, is less than $2k$, and

(c) for every $1 \leq i \leq t$, there exists a set M_i consisting of exactly $10^3k^2 \log n$ independent edges, each of which has exactly one endpoint in H_i.

Indeed, for showing (a), we note that if there are at least $4k$ independent edges between H_i to H_j, using the pigeonhole principle, at least k of them are between the same part of H_i (say S_i) and the same part of H_j (say S_j). Therefore, the graph obtained by joining H_i to H_j with this set of at least k edges is a k-connected (by Lemma 4.2.3) and bipartite graph that is larger than H_i; however, this is contrary to the maximality of H_i.

For showing (b), note that if there are at least $2k$ between v and H_i then there are at least k edges incident with v touch the same part of H_i, and let F be a set of k such edges. Second, we mention that joining a vertex of degree at least k to a k-connected graph trivially yields a k-connected graph. Next, since all the edges in F are touching the same part, the graph obtained by adding v to $V(H_i)$ and F to $E(H_i)$, will also be bipartite. This contradicts the maximality of H_i.

For (c), note first that since H_1 is not spanning, using (b) we conclude that in the construction of the bipartite subgraphs $H_1, \ldots, H_t$ in the process above,

$$\delta(G_2) \geq 10^{10}k^3 \log n - 2k \geq 8000k^2 \log n.$$

Therefore, using Corollary 4.2.2, it follows that G_2 contains a bipartite subgraph of size at least $10^3k^2 \log n$

which is also k-connected.

We will make use of the following theorem proven independently in 1931 by both König [49] and Egerváry [24]:

Theorem 4.3.3 (König–Egerváry Theorem)**.** *If G is a bipartite graph, then the maximum size of a matching in G equals the minimum size of a vertex cover of G.*

Therefore, the process does not terminate at this point, and H_2 exists (that is, $t \geq 2$). It also follows that for each $1 \leq i \leq t$ we have $|V(G) \setminus V(H_i)| \geq 10^3 k^2 \log n$. Next, note that G is $10^{10} k^3 \log n$ connected, and that each H_i is of size at least $10^3 k^2 \log n$. For each i, consider the bipartite graph with parts $V(H_i)$ and $V(G) \setminus V(H_i)$ and with the edge-set consisting of all the edges of G which touch both of these parts. Using König's Theorem, it follows that if there is no such M_i of size $10^3 k^2 \log n$, then there exists a set of strictly fewer than $10^3 k^2 \log n$ vertices that touch all the edges in this bipartite graph (a vertex cover). By deleting these vertices, one can separate what is left from H_i and its complement, contrary to the fact that G is $10^{10} k^3 \log n$ connected.

In order to complete the proof, we wish to reach a contradiction by showing that one can either merge few members of $\{H_1, \dots, H_t\}$ with vertices of V_0 into a k-connected component or find a k-connected component of size at least $10^3 k^2 \log n$ which is contained in V_0. In order to do so, we define an auxiliary digraph, using a special subgraph $G' \subseteq G$, and use Lemmas 4.3.2 and 4.2.4 to achieve the desired contradiction. We first describe how to find G'.

First, we partition V_0 into two sets, say A and B, where

$$A = \left\{ v \in V_0 : d_G\left(v, \bigcup_{i=1}^{t} V(H_i)\right) \geq 10^4 k^3 \log n \right\},$$

and observe that, using (b), since $A \subseteq V_0$, any vertex $a \in A$ must send edges to more than

$$10^4 k^3 \log n/(2k) = 5000 k^2 \log n$$

distinct elements in $X := \{H_1, \dots, H_t, v_1, \dots, v_s\}$. For each $1 \leq i \leq t$, let M_i be a set as described in (c). Observe that, using (b), each such M_i touches more than

$$10^3 k^2 \log n/(4k) = 250 k \log n$$

distinct elements of $X \setminus \{H_i\}$. Let $M_i' \subseteq M_i$ be a subset of size exactly $250 k \log n$ such that each pair of edges in M_i' touches two distinct elements of $X \setminus \{H_i\}$, which of course are distinct from G_i. Recall that $H_i = (S_i \cup T_i, E_i)$ for every $1 \le i \le t$.

For $Y := \{S_1, \ldots, S_t, T_1, \ldots, T_t, v_1, \ldots, v_s\}$, let

$$\Phi : Y \to \{L, R\}$$

be a mapping, generated according to the following random process:

Let $X_1, \ldots, X_t, Y_1, \ldots, Y_s \sim \text{Bernoulli}(1/2)$ be mutually independent random variables. For each $1 \le i \le t$, if $X_i = 1$, then let $\Phi(S_i) = L$ and $\Phi(T_i) = R$. Otherwise, let $\Phi(S_i) = R$ and $\Phi(T_i) = L$. For every $1 \le j \le s$, if $Y_j = 1$, then let $\Phi(v_j) = L$, and otherwise $\Phi(v_j) = R$. Now, delete all of the edges between two distinct elements of Y which receive the same label according to Φ.

Finally, define G' as the spanning bipartite graph of G obtained by deleting all of the edges within A and for distinct i and j, the edges between H_i and H_j which are not contained in $M_i' \cup M_j'$.

Recall by construction, using Φ we generated labels at random; therefore, by using Chernoff bounds (for instance see [4]), one can easily check that with high probability the following hold:

(i) For every $1 \le i \le t$, each set $M_i' \cap E(G')$ touches at least (say) $120k \log n$ other elements of X, and

(ii) for each $b \in B$, the degree of b into $A \cup B$ is at least (say) $d_{G'}(b, A \cup B) \ge 10^5 k^3 \log n$, and

(iii) for each vertex $a \in A$, there exist edges between a and $\cup_{i=1}^t V(H_i)$ that touch at least (say) $2000k^2 \log n$ distinct members of $\{H_1, \ldots, H_t\}$.

Note that here we relied on the luxury of losing the $\log n$ factor for using Chernoff bounds, but it seems like we could easily handle this "cleaning process" completely by hand.

Now we are ready to define our auxiliary digraph D. To this end, we first orient edges (again, bidirectional edges are allowed, and un-oriented edges are considered as bidirectional) of G' in the following way:

For every $1 \le i \le t$, we orient all of the edges in $E(G') \cap M_i'$ out of H_i. We orient all of the edges between A and $\cup_{i=1}^t V(H_i)$ out of A. We orient edges between B and $\cup_{i=1}^t V(H_i)$ arbitrarily, and we orient the remaining edges within $A \cup B$ in both directions.

Now, we define D to be the digraph with vertex set $V(D) = X$, and $\overrightarrow{xy} \in E(D)$ if and only if there exists an edge between x and y in G' which is oriented from x to y.

In order to complete the proof, we first note that with high probability D is a digraph on at most n vertices with out-degree $\delta^+(D) > (k-1)\lceil \log n \rceil$. This follows immediately from Properties (i)-(iii) as well as the way we oriented the edges. Therefore, one can apply Lemma 4.3.2 to find a sub-digraph $D' \subseteq D$ such that

1. For all $v \in V(D')$ we have $d^+_{D'}(v) \geq d^+_D(v) - (k-1)\lceil \log n \rceil$, and

2. $\kappa(U(D')) \geq k$.

In fact, with high probability, $\delta^+(D) \geq 120k \log n \geq k + (k-1)\lceil \log n \rceil$. Note that by construction, every pair of edges which are oriented out of some H_i must be independent and go to different components. Using Property 1. above combined with the fact that $\delta^+(D') \geq \delta^+(D) - (k-1)\lceil \log n \rceil \geq k$, we may conclude that the subgraph $G'' \subseteq G'$ induced by the union of all the components in $V(D')$ satisfies $G'' \in \mathcal{F}_{U(D')}(V(D'))$. Applying Lemma 4.2.4 with $X = V(D')$ and $R = U(D')$, it follows that $\kappa(G'') \geq k$.

In order to obtain the desired contradiction, we consider the following two cases:

Case 1: $V(G'')$ contains $V(H_i)$ for some i. We note that this case is actually impossible because it would contradict the maximality of H_i for the minimal index i such that $V(H_i) \subseteq V(G'')$.

Case 2: $V(G'') \subseteq A \cup B$. We note that in this case, there must be at least one vertex $b \in B \cap V(G'')$. Indeed, G'' is k-connected, and there are no edges within A. Now, it follows from Properties 1. and (ii) above that

$$d^+_{D'}(b) \geq d^+_D(b) - (k-1)\lceil \log n \rceil \geq 10^4 k^3 \log n.$$

Thus, it follows that $|V(G'')| \geq 10^4 k^3 \log n$. Combining this observation with the facts that G'' is k-connected and $V(G'') \subseteq A \cup B$, we obtain a contradiction. This case can not arise at all because of the fact that G'' should have been included as one of the bipartite subgraphs $\{H_1, \ldots, H_t\}$.

This completes the proof. □

4.4 Further Directions and Open Questions

Building on the ideas from Chapter 3 of this thesis a natural question to explore is: what can be said about random $f(k)$-regular graphs?

As shown in [14] not only is a random $f(k)$-regular graph is typically $f(k)$-edge-connected but a random $f(k)$-regular graph is typically $f(k)$-vertex-connected as well. Perhaps translating this problem to the setting of random regular graphs and using the Small Subgraph Conditioning Method could help us to understand in general what $f(k)$ might look like. More precisely, the following question is of interest:

Question 4.4.1. *For all k, does there exists a function $f(k)$ such that an $f(k)$-regular random graph G has a spanning, bipartite subgraph $H \subseteq G$ such that $\kappa(H) \geq k$ asymptotically almost surely?*

Perhaps in this context eliminating the $\log n$ factor would be possible.

The following appears as Conjecture 14 in Thomassen's [65] survey paper:

Conjecture 4.4.2. *For all k, there exists a function $g(k)$ such that for all digraphs D, if the vertex-connectivity*

$\kappa(D) \geq g(k)$, then there exists a (not necessarily spanning) k-vertex-connected bipartite sub-digraph.

This conjecture is still open, and it is a natural version of Conjecture 7 but for digraphs. Given that a key idea in our proof was to construct a special auxiliary digraph, perhaps similar methods could be applied here.

References

[1] N. Alon, J. Balogh, R. Morris, and W. Samotij, Counting sum-free sets in Abelian groups, *Israel J. Math.* **199** (2014), 309–344.

[2] N. Alon and F. R. K. Chung, Explicit constructions of linear sized tolerant networks, *Discrete Math.* **2** (1988), 15–19.

[3] N. Alon and P. Prałat, Modular orientations of random and quasi-random regular graphs, *Combin. Probab. Comput.*, **20** (2011), 321–329.

[4] N. Alon and J. H. Spencer, *The probabilistic method*, second edition, Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience, New York, 2000.

[5] R. Ahlswede and L. H. Khachatrian, The Complete Nontrivial-Intersection Theorem for Systems of Finite Sets, *J. Comb. Theory A*, **76** (1996), 121–138.

[6] L. Babai and P. Frankl, *Linear algebra methods in combinatorics with applications to geometry and computer science*, Department of Computer Science, The University of Chicago (1992).

[7] J. Balogh, T. Bohman, and D. Mubayi, Erdős–Ko–Rado in Random Hypergraphs, *Comb. Probab. Comput.*, **18** (2009), no. 5, 629–646.

[8] J. Balogh, B. Bollobás, and M. Simonovits, The typical structure of graphs without given excluded subgraphs, *Random Structures and Algorithms*, 34, no 3, 305–318.

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

[10] J. Balogh, R. Morris, and W. Samotij, Independent sets in hypergraphs, *J. Am. Math. Soc.*, 28 (2015), no. 3, 669–709.

[11] J. Barát and C. Thomassen, Claw-decompositions and Tutte-orientations, *J. Graph Theory*, **52** (2006), 135–146.

[12] J. Bensmail, A. Harutyunyan, T.-N. Le, M. Merker, and S. Thomassé, A proof of the Barat-Thomassen conjecture.

[13] A. Blokhuis, A. E. Brouwer, A. Chowdhury, P. Frankl, T. Mussche, B. Patkós, and T. Szőnyi, A Hilton–Milner theorem for vector spaces, *Electron. J. Combin.* **17** (2010), no. 1, Research Paper 71, 12.

[14] B. Bollobás. *Random Graphs*, Second Edition, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge (UK), 2001.

[15] B. Bollobás, On generalized graphs, *Acta Math. Acad. Sci. Hungar* **16** (1965), 447–452.

[16] B. Bollobás, A probabilistic proof of an asymptotic formula for the number of labelled regular graphs, *European J. Combin.*, **1** (1980), 311–316.

[17] P. J. Cameron and C. Y. Ku, Intersecting Families of Permutations, *European Journal of Combinatorics*, 24 (2003), 881-890.

[18] C. N. da Silva and R. Dahab, Tutte's 3-flow conjecture and matchings in bipartite graphs, *Ars Combin.*, 76 (2005), 83–95.

[19] R. Dedekind, Ueber Gruppen, deren sämmtliche Theiler Normaltheiler sind, *Math. Ann.*, Volume 48., No. 4 (1897), 548–561.

[20] M. Delcourt and A. Ferber, On a Conjecture of Thomassen, *Electronic Journal of Combinatorics*, Volume 22, Paper 3.2 (2015), 1–8.

[21] M. Delcourt and L. Postle, Random 4-regular graphs have S_3-decompositions asymptotically almost surely, arxiv:1610.01113 (2016).

[22] M. Deza and P. Frankl, On the maximum number of permutations with given maximal or minimal distance, *J. Comb. Theory A* **22** (1977), 352–360.

[23] R. Diestel, Graph Theory, Graduate Texts in Mathematics, Vol. 173, Springer, Berlin, (2000).

[24] E. Egerváry, On combinatorial properties of matrices (Hungarian with German summary), *Math. Lapok.*, **38** (1931), 16–28.

[25] D. Ellis, Stability for t-intersecting families of permutations, *J. Comb. Theory A* **118** (2011), 208–227.

[26] D. Ellis, Setwise intersecting families of permutations, *J. Comb. Theory A* **119** (2012), no. 4, 825–849.

[27] D. Ellis, E. Friedgut, and H. Pilpel, Intersecting families of permutations, *J. Am. Math. Soc.* **24** (2011), no. 3, 649–682.

[28] P. Erdős, P. Frankl, V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponent, *Graphs and Combin. 2* (2) (1986) 113–121.

[29] P. Erdős, D. J. Kleitman, and B. L. Rothschild, Asymptotic enumeration of K_n-free graphs, *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973), Tomo II* 19–27. Atti dei Convegni Lincei, 17, Accad. Naz. Lincei, Rome.

[30] P. Erdős, C. Ko, and R. Rado, Intersection theorems for systems of finite sets, *Q. J. Math.* **12.1** (1961), 313–320.

[31] P. Erdős and M. Simonovits, A limit theorem in graph theory, *Studia Sci. Math. Hungar.* **1** (1966), 51–57.

[32] P. Erdős and A. H. Stone, On the structure of linear graphs, *Bull. Amer. Math. Soc.* **52** (1946), 1087–1091.

[33] P. Frankl, The Erds-Ko-Rado theorem is true for $n = ckt$, *Combinatorics, Proc. Fifth Hungarian Coll. Combinatorics*, Keszthely, 1976, North-Holland, Amsterdam, (1978), 365–375.

[34] P. Frankl, An extremal problem for two families of sets, *Eur. J. Combin.* **3** (1982), 125–127.

[35] P. Frankl and A. Kupavskii, Counting intersecting and pairs of cross-intersecting families, arxiv:1701.04110 (2017).

[36] Z. Füredi, Geometrical solution of an intersection problem for two hypergraphs, *Eur. J. Combin.* **5** (1984), 133–136.

[37] M. Gauy, H. Hàn, and I. Oliveira, Erdős–Ko–Rado for random hypergraphs: asymptotic bounds and stability, preprint.

[38] H. Grötzsch, Herbert, Zur Theorie der diskreten Gebilde, VII: Ein Dreifarbensatz für dreikreisfreie Netze auf der Kugel, *Wiss. Z. Martin-Luther-U., Halle-Wittenberg, Math.-Nat. Reihe*, 8 (1959), 109–120.

[39] C. Godsil and K. Meagher, Erdős–Ko–Rado Theorems: Algebraic Approaches, Cambridge Studies in Advanced Mathematics. 1st Edition, (2016).

[40] A. Hamm and J. Kahn, On Erdős–Ko–Rado for random hypergraphs II, arXiv:1406.5793 (2014).

[41] A. J. Hilton and E. C. Milner, Some intersection theorems for systems of finite sets, *Q. J. Math.* **18.1** (1967), 369–384.

[42] W. N. Hsieh, Intersection theorems for systems of finite vector spaces, Discrete Math. **12** (1975), 1–16.

[43] F. Jaeger, On circular flows in graphs, *Finite and infinite sets*, Vol. I, II (Eger, 1981), vol. 37 of Colloq. Math. Soc. János Bolyai, North-Holland, Amsterdam, 1984, 391–402.

[44] F. Jaeger, Nowhere-zero flow problems, *Selected topics in graph theory*, **3**, Academic Press, San Diego, CA, 1988, 71–95.

[45] S. Janson, Random regular graphs: asymptotic distributions and contiguity, *Combin. Probab. Comput.*, **4** (1995), 369–405. 16

[46] J. Kahn, Personal communication.

[47] D. J. Kleitman and K. J. Winston, The asymptotic number of lattices, *Ann. Discrete Math.*, 6 (1980), 243–249, Combinatorial mathematics, optimal designs and their applications (Proc. Sympos. Combin. Math. and Optimal Design, Colorado State Univ., Fort Collins, Colo., 1978).

[48] Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers, *Random Struct. Algor.*, to appear.

[49] D. König, Graphen und Matrizen, *Math. Lapok.*, **38** (1931), 116–119.

[50] A. Kotzig, From the theory of finite regular graphs of degree three and four, *Casopis Pést. Mat.*, **82** (1957), 76–92 (in Slovak).

[51] H.-J. Lai, Mod $(2p+1)$-orientations and $K_{1,2p+1}$-decompositions, *SIAM J. Discrete Math.*, **21** (2007), 844–850.

[52] B. Larose and C. Malvenuto, Stable sets of maximal size in Kneser-type graphs, *European Journal of Combinatorics*, 25 (2004) 657–673.

[53] T. Łuczak, On triangle-free random graphs, *Random Structures Algorithms*, 16 (2000), 260–276.

[54] L. M. Lovász, C. Thomassen, Y. Wu, and C.-Q. Zhang, Nowhere-zero 3-flows and modulo k-orientations, *J. Combin. Theory Ser. B*, **103** (2013), 587–598.

[55] L. Lovász, Flats in matroids and geometric graphs, in *Combinatorial Surveys* (P. J. Cameron, ed.), Academic Press, New York (1977), 45–86.

[56] L. Lovász, On the Shannon capacity of a graph, *IEEE T. Inform. Theory* **25.1** (1979), 1–7.

[57] W. Mader, *Existenz n-fach zusammenhängender Teilgraphen in Graphen genügend grosser Kantendichte*, Abh. Math. Sem. Univ. Hamburg **37** (1972), pp. 86–97.

[58] W. Mantel, Problem 28, *Wiskundige Opgaven*, volume 10, (1907), 60–61.

[59] C. McDiarmid, Concentration, *Probabilistic methods for algorithmic discrete mathematics*, Springer Berlin Heidelberg, (1998), 195–248.

[60] P. Prałat and N. Wormald, Almost all 5-regular graphs have a 3-flow, arxiv:1503.03572 (2015).

[61] R.W. Robinson and N. C. Wormald, Almost all cubic graphs are Hamiltonian, *Random Structures Algorithms*, **3** (1992), 117–125.

[62] D. Saxton and A. Thomason, Hypergraph containers, *Inventiones Math.*, 201 (2015), 1–68.

[63] P. D. Seymour, Nowhere-zero flows. *Handbook of Combinatorics*, North-Holland, Amsterdam, (1995), 289–299.

[64] C. Thomassen, The weak 3-flow conjecture and the weak circular flow conjecture, *J. Combin. Theory Ser. B*, **102** (2012), 521–529.

[65] C. Thomassen, *Configurations in graphs of large minimum degree, connectivity, or chromatic number* in Combinatorial Mathematics: Proceedings of the Third International Conference (New York, 1985). Ann. New York Acad. Sci., Vol. 555, 1989, pp. 402–412.

[66] C. Thomassen, *Paths, circuits and subdivisions.* in *Selected topics in graph theory, 3*, Academic Press, San Diego, CA, 1988, pp. 97–131.

[67] P. Turán, Eine Extremalaufgabe aus der Graphentheorie, *Mat. Fiz. Lapok*, **48** (1941), 436–452.

[68] W. T. Tutte, A contribution to the theory of chromatic polynomials, *Canadian J. Math.*, **6** (1954), 80–91.

[69] W. T. Tutte, On the algebraic theory of graph colorings, *J. Combinatorial Theory*, **1** (1966), 15–50.

[70] D. B. West, *Introduction to graph theory.* Prentice Hall, Inc., Upper Saddle River, NJ, 1966.

[71] D. B. West, *The Art of Combinatorics, Volume I: Extremal Graph Theory.* Preliminary Version Spring 2013.

[72] D. B. West, *The Art of Combinatorics, Volume II: Structural Graph Theory.* Preliminary Version Fall 2014.

[73] R. M. Wilson, The exact bound in the Erdős–Ko–Rado theorem, *Combinatorica* **4** (1984), no. 2 − 3, 247–257.

[74] N.C. Wormald, Models of random regular graphs, *Surveys in combinatorics*, 1999 (Canterbury), vol. 267 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge, 1999, 239–298.