Methods in Extremal Combinatorics

Sam Spiro

August 21, 2024

Contents

Preface

This is a work in progress. There are typos, missing references, and so on scattered throughout. **Please let me know if you notices such errors, find anything confusing, or if you have any other suggestions!** If you prefer, you can let me know about any of this anonymously through this link.

The following is a set of lecture notes for a graduate level course in extremal combinatorics. These notes focus on standard methods that have been used to solve a large number of problems in extremal combinatorics. Throughout I assume basic knowledge of asymptotic analysis, probability theory, and linear algebra.

Due to the sheer scope of extremal combinatorics, there are many methods which I am not able to cover at all (and there is no topic which I am able to cover in complete depth). Below is a small list of methods and topics ***not*** covered by this text (which may be written up at some point), as well as some sources for thorough treatments of the topics.

- Extremal Combinatorics in general: see books of Lovasz [110] or Bollobás [25]; surveys by Simonovits and Szemerédi [136] and Füredi and Simonovits [73]; and online courses by Morris and Gowers [83].

- The Regularity Lemma: see the excellent book by Zhao [149] (as well as his corresponding video lectures).

- Additive combinatorics and discrete Fourier analysis: again Zhao [149] is a good introductory text, see also the book by Tao and Vu [141] and the online course by Prendiville.

- Discrete geometry: see the books by Sheffer [135] and Matoušek [114], as well as the online minicourse on finite geometry and Ramsey theory by Bishnoi.

- Statistical mechanics: see notes by Will Perkins.

- The discharging method: see the survey by Cranston and West [43].

- Absorption need to find a reference.

Major Updates

Here I whatever major additions/rewritings I have done since the last posting in case anyone wants to check out what's new.

- 8/21/24:
  - Added new chapters on Linear Programming and Homomorphism Counting.
  - Created a new part on Matchings in Hypergraphs, which includes a modified version of my previous writeup on the Rödl Nibble, as well as a new chapter on the Forbidden Submatching Method.
  - Added a new example Theorem 3.5 to the chapter on the Local Lemma.

Part I

Basic Probabilistic Methods

This part is based heavily off of the book by Alon and Spencer [8] (which goes into much more depth on the topic), as well as lecture notes by Verstraëte.

1 Introduction

One of the most exciting developments in extremal combinatorics over the past century has been the incorporation of ideas and tools from probability theory into solving combinatorial problems. The first such use was by Erdős who proved an exponential lower bound for Ramsey numbers. We recall that the *Ramsey number* $R(s,t)$ is the smallest integer N such that any 2-coloring of the edges of K_N contains a monochromatic clique.

Theorem 1.1 ([55])**.** *For all n, we have*

$$R(n,n) \geq (1+o(1))\frac{n}{e\sqrt{2}}2^{n/2}.$$

This is essentially the best known lower bound (though we prove a slightly stronger bound in Theorem 3.3). The best known upper bound is roughly 4^n, so there's still quite a gap!

For this proof and throughout the text, we make heavy use of the union bound: if A, B are events in a probability space, then $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$. Often we will use an equivalent version: $\Pr[\overline{A} \cap \overline{B}] \leq 1 - \Pr[A] - \Pr[B]$, which follows from De Morgan's laws.

Proof. Let G be a ***random*** coloring of K_N with N to be determined later[1]. That is, for each edge of K_N, we independently and uniformly choose the edge to be colored either red or blue. The key observation is that if $\Pr[G \text{ contains no monochromatic } K_n] > 0$, then there exists a coloring of K_N with no monochromatic K_n (since otherwise the probability would be zero), proving the desired lower bound.

If S is a set of n vertices, we let A_S be the event that G contains a monochromatic K_n on S. With this we have

$$\Pr[G \text{ contains a monochromatic } K_n] = \Pr\Big[\bigcup_{S \in \binom{[N]}{n}} A_S\Big] \leq \sum_{S \in \binom{[N]}{n}} \Pr[A_S] = \binom{N}{n} \cdot 2^{1-\binom{n}{2}}.$$

If this quantity is less than 1, then we can conclude that $\Pr[G \text{ contains no monochromatic } K_n] > 0$, so our goal is to choose N as large as possible so that this happens. By using the bound $\binom{N}{n} \leq (eN/n)^n$ (which we will use many times throughout the text), we see that it suffices to have[2]

$$1 > (eN/n)2^{1-\binom{n}{2}} = 2(eN/n2^{(n-1)/2})^n.$$

Solving this shows that the desired bound holds if $N < 2^{1/n} \cdot \frac{n}{e\sqrt{2}}2^{n/2}$, proving the result[3]. □

[1] When trying to prove results in extremal and probabilistic combinatorics, one often uses a method that depends on some parameter such as N or p. Typically it is best to proceed through the argument without deciding what N, p is ahead of time, and only in the end do you optimize your parameter to give you the best bounds possible.

[2] Finding the "right" way to bound expressions like this takes time and practice. A reasonable strategy for these sorts of problems is try and get all of the main terms to have the same form (e.g. x^n in this example). Much more about the art of asymptotic analysis can be found in the book Asymptopia by Spencer [138].

[3] In fact, a closer analysis of this proof shows that asymptotically, almost every coloring of K_N with $N = (2-\epsilon)^{n/2}$ contains no monochromatic K_n. Despite almost every coloring working, we know of no explicit coloring that gives more than a polynomial lower bound for $R(n,n)$. Thus the probabilistic method gives us a way to find the hay in the haystack.

The proof of Theorem 1.1 implicitly used the following general principle, which is at the heart of the probabilistic method.

(*) Let T be an object chosen randomly from a set $\mathcal{T}$ (in some way) and P some property that objects in $\mathcal{T}$ could have. If $\Pr[T \text{ has property } P] > 0$, then there exists some $T' \in \mathcal{T}$ with this property.

We now turn to another classical extremal problem with a slick probabilistic proof. Recall that $\alpha(G)$ denotes the largest independent set of a graph G, i.e. the largest set of vertices I such that there exist no edge contained in I.

Theorem 1.2 (Caro-Wei Bound). *Let G be an n-vertex graph with degrees $d_1, \ldots, d_n$. Then*

$$\alpha(G) \geq \sum \frac{1}{d_i + 1}.$$

Moreover, equality holds if and only if G is a disjoint union of cliques.

Here and throughout the text we make heavy use of the principle of linearity of expectation: for two (possibly dependent) real-valued random variables, we have $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Proof. For π a bijection from $V(G)$ to $[n]$, we define

$$I(\pi) = \{v \in V(G) : \pi(v) < \pi(u)\ \forall u \in N(v)\}.$$

That is, $I(\pi)$ is the set of vertices which are smaller than all of their neighbors under π. Observe that $I(\pi)$ is an independent set (if u, v are adjacent we must have, say $\pi(v) < \pi(u)$, in which case $u \notin I(\pi)$), so in particular $\alpha(G) \geq |I(\pi)|$ for all π.

Let π be a random bijection chosen uniformly amongst all bijections from $V(G)$ to $[n]$, and let 1_v be the indicator variable which is 1 if $v \in I(\pi)$ and 0 otherwise. Note that regardless of what π is, we have $\alpha(G) \geq |I(\pi)| = \sum 1_v$, so by linearity of expectation we have

$$\alpha(G) \geq \mathbb{E}[I(\pi)] = \sum \mathbb{E}[1_v] = \sum \Pr[1_v = 1]. \tag{1}$$

Observe that $1_v = 1$ if and only if $\pi(v) = \min_{u \in \{v\} \cup N(v)} \pi(u)$. Since π was chosen uniformly at random, each $u \in \{v\} \cup N(v)$ is equally likely to achieve this minimum, so $\Pr[1_v = 1] = \frac{1}{d(v)+1}$, and plugging this into (1) gives the result.

Note that equality holds in (1) if and only if $I(\pi)$ is an independent set of maximum size for all bijections π. It is not too difficult to show that this holds if and only if G is a disjoint union of cliques, and we leave this as an exercise to the reader. $\square$

Theorem 1.2 implies Turán's theorem, which is essentially the result that jump started the entire field of extremal combinatorics[1] (though the original proof was not probabilistic).

[1]The first theorem in extremal combinatorics is typically attributed to Mantel, which is the $r = 3$ case of Turán's Theorem. However, it wasn't until Turán's result 30 years later that the field really took off.

To state this result, we define $\mathrm{ex}(n, F)$ to be the largest number of edges that an n-vertex F-free graph can have[1] which is called the *Turán number* or *extremal number* of F. We define the *Turán graph* $T_r(n)$ to be the complete n-vertex r-partite graph with parts of sizes as equal as possible. We let $t_r(n) = e(T_r(n))$. For example, $T_2(n) = K_{\lfloor n/2\rfloor,\lceil n/2\rceil}$ and $t_2(n) = \lfloor n/2\rfloor \cdot \lceil n/2\rceil = \lfloor n^2/4\rfloor$. More generally we have

$$t_r(n) \le \binom{r}{2}(n/r)^2 = \frac{r-1}{r}\cdot\frac{n^2}{2} = \left(1-\frac{1}{r}\right)\frac{n^2}{2},$$

with equality holding if $r|n$ and otherwise $t_r(n)$ is the floor of this upper bound.

Corollary 1.3 (Turán's Theorem)**.** *For all $r \le n$ we have*

$$\mathrm{ex}(n, K_r) = t_{r-1}(n).$$

Moreover, $T_{r-1}(n)$ is the unique n-vertex K_r-free graph with $t_{r-1}(n)$ edges.

Proof. The lower bound $\mathrm{ex}(n, K_r) \ge t_{r-1}(n)$ follows by considering $T_{r-1}(n)$. Let G be an n-vertex K_r-free graph with degrees $d_1, \ldots, d_n$. Observe that the complement $\overline{G}$ contains no independent set of size r, so by Theorem 1.2 we have

$$r - 1 \ge \alpha(\overline{G}) \ge \sum \frac{1}{n - d_i}. \tag{2}$$

Observe that if x, y are positive numbers, then[2]

$$\frac{1}{x} + \frac{1}{y} \ge \frac{1}{\frac{1}{2}(x+y)} + \frac{1}{\frac{1}{2}(x+y)}$$

with equality holding if and only if $x = y$. In view of this inequality, we see that (2) is minimized when all of the d_i are as close together as possible. Because $\sum d_i = 2e(G)$, we have

$$r - 1 \ge n \cdot \frac{1}{n - 2e(G)/n} = \frac{n^2}{n^2 - e(G)} \implies e(G) \le \left(1 - \frac{1}{r-1}\right) n^2/2,$$

so $e(G) \le t_{r-1}(n)$ as desired. Moreover, to have equality, $\overline{G}$ must be a union of cliques with sizes as close as possible to each other, i.e. G must be a complete r-partite graph with parts having sizes as close as possible to each other, i.e. G must be the Turán graph. $\square$

In addition to using the probabilistic method to get an upper bound for $\mathrm{ex}(n, K_n)$ as in Corollary 1.3, one can also use it to give a general lower bound for $\mathrm{ex}(n, F)$.

Theorem 1.4. *Let F be a graph with v vertices and $e \ge 2$ edges. If $e \ge v$, then*

$$\mathrm{ex}(n, F) = \Omega_v(n^{2-\frac{v-2}{e-1}}).$$

[1] Throughout the text, a graph being F-free means that it contains no subgraph which is isomorphic to F (and we don't care whether this subgraph is induced or not).

[2] By multiplying both sides of the above expression by $xy(x+y)$, we see that this is equivalent to saying $y(x+y) + x(x+y) \ge 4xy$, which is equivalent to saying $x^2 - 2xy + y^2 = (x-y)^2 \ge 0$.

For this proof we use an object that is fundamental to probabilistic and extremal combinatorics. This is the *Erdös-Renyi random graph* $G_{n,p}$, which is the random graph on n vertices that contains each edge $e \in E(K_n)$ independently and with probability p. For example, $G_{n,1} = K_n$ and $G_{n,1/2}$ is equally likely to be any labeled graph on n vertices. The random graph is an incredibly fascinating object in its own right. We will not discuss it in too much depth in this text, see the book by Frieze and Karoński [71] for a thorough treatment of it.

Proof. Let $G_{n,p}$ be the random graph with p a quantity to be determined later. Let X denote the number of copies of F in $G_{n,p}$. For S a set of v vertices, let 1_S be the indicator variable which is 1 if S contains a copy of F in $G_{n,p}$ and which is 0 otherwise. With this,

$$\sum 1_S \le X \le v! \sum 1_S,$$

since each set of v vertices contains at most $v!$ copies of F. To have $1_S = 1$, we in particular need S to contain at least e edges, so

$$\Pr[1_S = 1] \le \sum_{k \ge e} \binom{\binom{v}{2}}{k} p^k (1-p)^{\binom{v}{2}-k} \le v^2 2^{v^2} p^e \le 4^{v^2} p^e.$$

In total this gives

$$\mathbb{E}[X] \le v! \binom{n}{v} \cdot 4^{v^2} p^e \le (4^v n)^v p^e.$$

Observe that when $p \gg n^{v/e}$, the calculation above suggests that $G_{n,p}$ will contain copies of F (at least in expectation), so $G_{n,p}$ will not work as an F-free graph for this range of p. However, we can get around this by using the following trick known as the method of alterations. Let G be any subgraph of $G_{n,p}$ obtained by deleting an edge from each copy of F in $G_{n,p}$. By definition G will be F-free. Moreover, the number of edges that G has is at least $e(G_{n,p}) - X$ since at most X of the original edges from $G_{n,p}$ are deleted. Using linearity of expectation gives

$$\mathbb{E}[e(G)] \ge \mathbb{E}[e(G_{n,p}) - X] \ge p\binom{n}{2} - (4^v n)^v p^e \ge \frac{1}{4} p n^2 - (4^v n)^v p^e. \tag{3}$$

At this point we want to choose p so that the above expression is roughly maximized. Intuitively this will happen when both terms on the rightside of (3) are roughly equal to each other, i.e. when $pn^2 \approx n^v p^e$. This suggests taking $p \approx n^{\frac{2-v}{e-1}}$. And indeed, after playing around for a bit, one sees that, for example, taking $p = \frac{1}{20 \cdot 16^v} n^{\frac{2-v}{e-1}}$ and plugging it into (3) gives[1] $\mathbb{E}[e(G)] \ge \frac{1}{160 \cdot 16^v} n^{2-\frac{2-v}{e-1}}$. Because G is a (random) F-free graph, by (*) there exists some deterministic graph G' which is F-free with this many edges, proving the result. $\square$

For many F, there are known constructions which give much better lower bounds for $\mathrm{ex}(n, F)$ than Theorem 1.4. However, this is the best known lower bound which works for arbitrary F.

The method used in this proof is known as the method of alterations. Typically this works by defining some initial random set A (e.g. a set of edges of a graph) which contains some bad

[1] Here we use $4^{v^2} \le 4^{ve}$ and that $e \ge 2$.

subsets B (e.g. subsets of edges forming a forbidden graph F). We then define a random set A' by deleting an element from each bad subset B, giving that $|A'| \geq |A| - |B|$ and that A' has no bad subsets. At this point we win provided

$$\mathbb{E}[|A'|] = \mathbb{E}[|A|] - \mathbb{E}[|B|]$$

is large. Typically the expectations $\mathbb{E}[|A|], \mathbb{E}[|B|]$ depend on some common parameter p, and we often optimize this expression by finding p such that $\mathbb{E}[|A|] \approx \mathbb{E}[|B|]$, and then ultimately choosing p to be a bit smaller than this so that, say, $\mathbb{E}[|B|] \leq \frac{1}{2}\mathbb{E}[|A|]$.

(**) The method of alterations detailed above is often very useful.

The last core tenant of the probabilistic method that we have implicitly used throughout this section is the following.

(***) If one is trying to find a nice object, one should always try and see how well a random object does (possibly after applying alterations).

For example, the most straightforward random coloring gave the bound of Theorem 1.1, and the random graph together with alterations gave Theorem 1.4.

Lastly, we note that in principle many of these results could be proven without needing to use probability. However, for certain problems a probabilistic perspective is genuinely useful since it is allows one to use powerful tools from probability theory (e.g. martingales and concentration inequalities). Even when it isn't strictly needed, probability often provides for a much clearer perspective on a problem.

2 Some Random Examples

This section consists of an assorted collection of examples which provides both practice with the general principles of the probabilistic method, as well as proofs of many fundamental results from extremal combinatorics.

2.1 Graphs with Small and Large Chromatic Numbers

We start with a very simple example that will be used throughout the text (often without reference).

Lemma 2.1. *If G is an n-vertex graph, then there exists a bipartite subgraph $G' \subseteq G$ such that $e(G') \geq \frac{1}{2}e(G)$. Moreover, we can choose G' such that its partition classes U, V have sizes $\lfloor n/2 \rfloor, \lceil n/2 \rceil$.*

Given this lemma, if you want to prove a statement of the form "any graph G with $\Omega(m)$ edges has some monotone graph property", then you only need to consider graphs which are (balanced) bipartite.

Proof. The first part is very easy: let $U \subseteq V(G)$ be obtained by including each vertex independently and with probability $\frac{1}{2}$, and let $V = V(G) \setminus U$. Let G' be the graph which consists of every edge $e \in E(G)$ with one vertex in U and one vertex in V. It is easy to check that $\mathbb{E}[e(G')] = \frac{1}{2}e(G)$, so such a (bipartite) subgraph exists.

The second part is conceptually easy but computationally a little tedious. Let $U \subseteq V(G)$ be a set of size $\lfloor n/2 \rfloor$ chosen uniformly at random and let $V = V(G) \setminus U$. Let G' be the graph which consists of every edge $e \in E(G)$ with one vertex in U and one vertex in V. Observe that the probability that a given edge $xy \in E(G)$ is in G' is exactly

$$1 - \frac{\lfloor n/2 \rfloor \cdot (\lfloor n/2 \rfloor - 1)}{n(n-1)} - \frac{\lceil n/2 \rceil \cdot (\lceil n/2 \rceil - 1)}{n(n-1)} \geq \frac{1}{2},$$

with the last step following from a case analysis based on whether n is even or odd. Thus in expectation G' has at least $\frac{1}{2}e(G)$ edges, so such a balanced bipartite subgraph of G must exist. $\square$

A graph G is said to have *girth* ℓ if its smallest cycle is of size ℓ, and we say that it has infinite girth if G has no cycles. Observe that graphs of large girth locally look like a tree, i.e. if you pick any vertex v, then the graph induced by every vertex within distance ℓ of v is a tree. In particular, "locally" graphs of large girth can be properly colored using few colors, but does this necessarily hold globally as well? That is, does there exist graphs with girth at least ℓ and chromatic number at least k for all ℓ, k? A clever (random) argument of Erdős shows that such a graph does indeed exist.

Theorem 2.2 (Erdős)**.** *For all ℓ, k there exist graphs of girth at least ℓ and chromatic number at least k.*

For this proof we use Markov's inequality: if X is a non-negative real-valued random variable, then $\Pr[X \geq x] \leq \mathbb{E}[X]/x$ for $x > 0$.

Proof. Consider $G_{n,p}$ with n, p to be determined later. Let $X_{\leq \ell}$ denote the number of cycles in $G_{n,p}$ of size at most ℓ. Linearity of expectation gives

$$\mathbb{E}[X_{\leq \ell}] \leq \sum_{t=3}^{\ell} n^t \cdot p^t \leq \ell (pn)^\ell.$$

Thus if we wanted $G_{n,p}$ to have girth smaller than ℓ with high probability, by Markov's inequality it would suffice to take $p \ll n^{-1}$. Unfortunately this naive approach is too weak since in this case $G_{n,p}$ will have very small chromatic number. To get around this, we will take p slightly larger than n^{-1} and then use alterations to delete a vertex from every small cycle of $G_{n,p}$. With some foresight[1] we will take $p = n^{-1+1/2\ell}$. With this we see that

$$\Pr[X_{\leq \ell} \geq n/2] \leq \mathbb{E}[X_{\leq \ell}]/(n/2) \leq 2\ell n^{-1/2}. \tag{4}$$

We now turn to the chromatic number of $G_{n,p}$, which is a slightly trickier quantity to get a handle on. To do this we use the inequality $\chi(G) \geq |V(G)|/\alpha(G)$, which follows from the fact that a k-coloring of G is a partition of $V(G)$ into independent sets. Thus for $G_{n,p}$ to have large chromatic number, it suffices to show that all of its independent sets are small . For m an integer we let Y_m be the number of independent sets of size m in $G_{n,p}$. Using linearity of expectation and $(1-x) \leq e^{-x}$ gives for $m \geq 2$

$$\mathbb{E}[Y_m] = \binom{n}{m} \cdot (1-p)^{\binom{m}{2}} \leq n^m \cdot \left(e^{-p(m-1)/2}\right)^m \leq \left(ne^{-pm/4}\right)^m.$$

By Markov's inequality and our choice of $p = n^{-1+1/2\ell}$, we find for $m = n/2k$ and n sufficiently large in terms of k, ℓ that

$$\Pr[Y_{n/2k} \geq 1] \leq (ne^{-n^{1/2\ell}/8k})^m < \frac{1}{2}. \tag{5}$$

By combining (4) and (5), we see for n sufficiently large that $X_{\leq \ell} < n/2$ and $Y_{n/2k} = 0$ both occur with positive probability, i.e. there exists a graph G such that both of these events occur. Let G' be G after deleting a vertex from each cycle of length at most ℓ in G. This deletes at most half the vertices of G by assumption of $X_{\leq \ell}$, and we have $\alpha(G') \leq \alpha(G) \leq n/2k$. Thus

$$\chi(G') \geq |V(G')|/\alpha(G') \geq k,$$

proving the result. ∎

2.2 Random Permutations and Extremal Set Theory

In this subsection, we use random permutations (similar to the proof of Theorem 1.2) to prove two famous results from extremal set theory, which is roughly speaking the study of

[1] The exact choice of p doesn't matter here, the important thing is to take $p = n^{-1+\alpha}$ with $0 < \alpha < 1/\ell$.

extremal problems for hypergraphs. We only scratch the surface of this topic, see Frankl and Tokushige [66] for a more thorough treatment.

We start with the most fundamental theorem in extremal set theory: the Erdős-Ko-Rado theorem.

Theorem 2.3 (Erdős-Ko-Rado Theorem)**.** *Let $\mathcal{F} \subseteq \binom{[n]}{k}$ be an intersecting family, i.e. $F \cap F' \neq \emptyset$ for any $F, F' \in \mathcal{F}$. If $n \geq 2k$, then*

$$|\mathcal{F}| \leq \binom{n-1}{k-1}.$$

This bound is sharp by taking $\mathcal{F}$ to consist of every set containing the element 1 (and in fact, up to isomorphism this is the unique extremal construction when $n > 2k$). Note that if $n < 2k$, then $\mathcal{F} = \binom{[n]}{k}$ is an intersecting family, so we need $n \geq 2k$ for us to be able to prove a non-trivial bound.

Proof. The proof uses what is known as Katona's circle method, which involves choosing a random cyclic ordering $\pi : [n] \to \mathbb{Z}_n$, where $\mathbb{Z}_n$ is the integers mod n. Given such a π and a set $A \in \mathcal{F}$, we let 1_A be the indicator variable with $1_A = 1$ if $A = \{\pi(i), \pi(i)+1, \ldots, \pi(i)+k-1\}$ for some $i \in [n]$. We claim that $1_A = 1$ for at most k sets A.

Indeed, if $1_A = 0$ for all A then there is nothing to prove, so assume $1_A = 1$ for some A, say with $A = \{\pi(i), \pi(i)+1, \ldots, \pi(i)+k-1\}$. Let $S_j = \{\pi(i)+j, \pi(i)+j+1, \ldots, \pi(i)+j+k-1\}$, and observe that if $B \in \mathcal{F}$ has $1_B = 1$, then we must have[1] $B = S_j$ for some $-k < j < k$. Moreover, for each pair $\{S_{-k+\ell}, S_\ell\}$ with $0 \leq \ell < k$, at most one $B \in \mathcal{F}$ is equal to one of these sets since $S_{-k+\ell}, S_\ell$ are disjoint, so in total we conclude that $1_A = 1$ for at most k different $A \in \mathcal{F}$.

Observe[2] that $\Pr[1_A = 1] = n\binom{n}{k}^{-1}$, and this together with the claim above implies

$$k \geq \mathbb{E}[\sum_{A \in \mathcal{F}} 1_A] = \sum_{A \in \mathcal{F}} \Pr[1_A = 1] = |\mathcal{F}| \cdot n\binom{n}{k}^{-1},$$

and rearranging gives the desired bound. ☐

There are many, many proofs of the Erdős-Ko-Rado theorem, as well as many generalizations and applications. Again, we refer the reader to [66] for more on this. Our second result related to extremal set theory is the following.

Theorem 2.4 (Bollobás Set Pairs Inequality)**.** *Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ and $\mathcal{B} = \{B_1, \ldots, B_m\}$ be set systems such that $A_i \cap B_i = \emptyset$ for all i and $A_i \cap B_j \neq \emptyset$ for all $i \neq j$. Then*

$$\sum_{i=1}^{m} \binom{|A_i| + |B_i|}{|A_i|}^{-1} \leq 1$$

[1] Here we use that each $B \in \mathcal{F}$ intersects A and that $n \geq 2k$ implies S_k is disjoint from A

[2] This follows because for any cyclic ordering π there are exactly n sets S which have $1_S = 1$

Pairs of families as in Theorem 2.4 are called *cross-intersecting*.

Proof. Let π be a random permutation of the underlying ground set (the size of which is irrelevant for the conclusion/proof). Let 1_i be the indicator variable with $1_i = 1$ if $\pi(x) < \pi(y)$ for all $x \in A_i$ and $y \in B_i$. That is, 1_i is the indicator for the event that A_i appears completely before B_i under π. A simple counting argument shows that $\Pr[1_i = 1] = \binom{|A_i|+|B_i|}{|A_i|}^{-1}$ (where here we implicitly use that $A_i \cap B_i = \emptyset$, as otherwise $\Pr[1_i = 1] = 0$).

We claim that there is at most one i such that $1_i = 1$. Indeed, say $1_i = 1$. Then for any $j \neq i$, by hypothesis there is some $x \in A_j \cap B_i \subseteq A_j$ and $y \in A_i \cap B_j \subseteq B_j$, and since $1_i = 1$, we have $\pi(x) > \pi(y)$. Thus $1_j = 0$ for all $j \neq i$. With this claim we have

$$1 \geq \mathbb{E}\Big[\sum_i 1_i\Big] = \sum \Pr[1_i = 1] = \sum \binom{|A_i| + |B_i|}{|A_i|}^{-1}.$$

□

Theorem 2.4 has many applications. One such application involves *antichains*, which are collections of sets $\mathcal{F}$ such that there exist no distinct $A, B \in \mathcal{F}$ with $A \subseteq B$.

Corollary 2.5 (LYM Inequality)**.** *If $\mathcal{F} \subseteq 2^{[n]}$ is an antichain, then*

$$\sum_{A \in \mathcal{F}} \binom{n}{|A|}^{-1} \leq 1.$$

Proof. Let $\mathcal{F} = \{A_1, \ldots, A_m\}$ and define $B_i = [n] \setminus A_i$. It is not difficult to check that since $\mathcal{F}$ is an antichain, $A_i \cap B_j = \emptyset$ if and only if $i = j$. The bound then follows from Theorem 2.4. □

We note that the proof of Corollary 2.5 is a nice simplification of the proof of Theorem 2.4: now $1_i = 1$ if and only if $A_i = \{\pi(1), \ldots, \pi(|A_i|)\}$.

Corollary 2.6 (Sperner's Theorem)**.** *If $\mathcal{F} \subseteq 2^{[n]}$ is an antichain, then*

$$|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

This result is sharp, as can be seen by taking $\mathcal{F} = \binom{[n]}{\lfloor n/2 \rfloor}$ or $\binom{[n]}{\lceil n/2 \rceil}$.

Proof. We have $\binom{n}{k} \leq \binom{n}{\lfloor n/2 \rfloor}$ for all k, so by the LYM inequality

$$1 \geq \sum_{A \in \mathcal{F}} \binom{n}{|A|}^{-1} \geq |\mathcal{F}| \binom{n}{\lfloor n/2 \rfloor}^{-1},$$

and moving things around gives the desired result. □

2.3 The Crossing Lemma and Incidence Geometry

Our final result concerns drawings of graphs. Without being too precise with our definitions, we define the *crossing number* of a graph G to be the minimum number of crossings that an embedding $\phi(G)$ in the plane will have. For example, a graph is planar if and only if $cr(G) = 0$.

Lemma 2.7. *If G is an n-vertex graph with m edges, then $cr(G) \geq m - 3n$.*

Sketch of Proof. Let $\phi(G)$ be an embedding of G with $cr(G)$ crossings. By deleting an edge from each crossing, we obtain a planar graph G' with n vertices and at least $m - cr(G)$ edges. A simple consequence of Euler's formula shows that this means $m - cr(G) \leq 3n$, giving the result. □

We will use the probabilistic method to "amplify" the elementary bound of Lemma 2.7 and give a bound that is effective for dense graphs.

Lemma 2.8 (Crossing Lemma)**.** *If G is an n-vertex graph with $m \geq 4n$ edges, then*

$$cr(G) \geq \frac{m^3}{64n^2}.$$

Proof. Let $\phi(G)$ be an embedding of G which has $cr(G)$ crossings. Let $V_p \subseteq V(G)$ be obtained by keeping each vertex of $V(G)$ independently and with probability p, and let $G_p = G[V_p]$. Observe that there is a natural embedding of G_p, namely the restriction of ϕ to G_p.

Let X denote the number of crossings in $\phi(G_p)$, and note that $\mathbb{E}[X] = p^4 cr(G)$ since a crossing survives if and only if all four of its relevant vertices lie in V_p. Using Lemma 2.7, we see that

$$p^4 cr(G) = \mathbb{E}[X] \geq \mathbb{E}[e(G') - 3|V_p|] = p^2 m - 3pn \implies cr(G) \geq p^{-2}m - 3p^{-3}n.$$

This lower bound will roughly be optimized when $p^{-2}m = p^{-3}n$, i.e. when $p = n/m$. More precisely, taking $p = 4n/m$ gives the desired bound. However, implicitly this argument requires that $0 \leq p \leq 1$, i.e. that $m \geq 4n$, and this holds by hypothesis. □

In addition to being interesting in its own right, the crossing lemma gives a short proof of a fundamental result in incidence geometry.

Theorem 2.9 (Szemeredi-Trotter Theorem)**.** *Let $\mathcal{P}$ be a set of n points and $\mathcal{L}$ a set of m lines in the plane, and let $\mathcal{I} \subseteq \mathcal{P} \times \mathcal{L}$ denote their set of incidences, i.e. pairs (p, ℓ) with $p \in \ell$. Then*

$$|\mathcal{I}| = O(m^{2/3}n^{2/3} + m + n).$$

This bound is essentially best possible, though we omit the details of the (not too difficult) construction.

Proof (due to Székely). Without loss of generality, we can assume every point and line is in at least one incidence (otherwise we can delete these points/lines). Let G be the graph on $\mathcal{P}$ which makes two points p_1, p_2 adjacent if there exists a line $\ell \ni p_1, p_2$ such that there is no third point

q on the line segment p_1p_2. In other words, G is the graph obtained by drawing the points and lines on the plane, and then erasing the rays of lines which go off to infinity.

If $i(\ell)$ denotes the number of points incident to ℓ, then it is not difficult to see that $e(G) = \sum i(\ell) - 1 = |\mathcal{I}| - m$, where here we implicitly used that $i(\ell) \geq 1$ for all ℓ. If $|\mathcal{I}| \leq 2m$, then in particular $|\mathcal{I}| = O(m)$ and the result follows, so we can assume $e(G) \geq \frac{1}{2}|\mathcal{I}|$, and similarly we can assume $|\mathcal{I}| \geq 8m$ and hence $e(G) \geq 4n$. Thus by the crossing lemma we have

$$cr(G) \geq \frac{|\mathcal{I}|^3}{2^9 n^2}.$$

The critical observation is that $cr(G) \leq \binom{m}{2}$ since each crossing corresponds to two lines of $\mathcal{L}$ intersecting. Plugging this into the expression above gives the desired result. $\square$

As a brief aside, we note that this idea of taking a weak result (Lemma 2.7) and amplifying it to a stronger result (Lemma 2.8) shows up in many other places in extremal combinatorics. For example, it is easy to prove a weak version of the Szemeredi-Trotter theorem with a bound of roughly $O(mn^{1/2} + n)$ by observing that there exist no points p_1, p_2 and ℓ_1, ℓ_2 such that all of the incidences (p_i, ℓ_j) are present, i.e. the "incidence graph" on $\mathcal{P} \cup \mathcal{L}$ contains no C_4. One can then use the method of polynomial partitioning to dissect $\mathbb{R}^2$ into small regions where this bound is effective. For much more on incidence geometry and polynomial partitioning, we refer the reader to the excellent book by Sheffer [135].

3 The Lovász Local Lemma

We say that an event A_i is mutually independent of a set of events $\{A_j : j \in J\}$ if for any $J' \subseteq J$, we have $\Pr[A_i \cap \bigcap_{j\in J'} A_j] = \Pr[A_i] \cdot \Pr[\bigcap_{j\in J'} A_j]$. We say that $A_1, \ldots, A_n$ are mutually independent events if A_i is mutually independent of $\{A_j : j \in [n] \setminus \{i\}\}$ for all i. Note that in this case we have $\Pr[\bigcap A_i] = \prod \Pr[A_i]$. In this section we consider a result which roughly says that if the A_i's are "almost independent", then we have $\Pr[\bigcap A_i] \approx \prod \Pr[A_i]$.

Theorem 3.1. *[Lovász Local Lemma] Let $A_1, \ldots, A_n$ be events and let $D_1, D_2, \ldots, D_n \subseteq [n]$ be such that A_i is mutually independent of $\{A_j : j \notin D_i \cup \{i\}\}$ for all i. If there exist real numbers $\gamma_i \in [0, 1)$ such that $\Pr[A_i] \leq \gamma_i \prod_{j\in D_i}(1 - \gamma_j)$ for all i, then*

$$\Pr[\bigcap \overline{A_i}] \geq \prod(1 - \gamma_i) > 0.$$

This result is often just referred to as "the local lemma". Note that if the A_i were all mutually independent, then we could take $D_i = \emptyset$ and $\gamma_i = \Pr[A_i]$ for all i and conclude from the local lemma that $\Pr[\bigcap \overline{A_i}] \geq \prod \Pr[\overline{A_i}]$.

Proof. We claim that for all i and $S \subseteq [n]$, we have

$$\Pr[\overline{A_i} | \bigcap_{j\in S} \overline{A_j}] \geq 1 - \gamma_i.$$

This will give the result since then

$$\Pr[\bigcap_i \overline{A_i}] = \prod_i \Pr[\overline{A_i} | \bigcap_{j\in[i-1]} \overline{A_j}] \geq \prod_i (1 - \gamma_i).$$

We prove this claim by induction[1] on $|S|$. The base case $|S| = 0$ is equivalent to saying $\Pr[A_i] \leq \gamma_i$ for all i, and this follows from $\Pr[A_i] \leq \gamma_i \prod_{j\in D_i}(1-\gamma_j) \leq \gamma_i$. Now consider any set S, and in particular assume we have proven the result for all $S' \subsetneq S$. If $i \in S$ then the result is trivial, so we can assume $i \notin S$. Observe that

$$\begin{aligned}\Pr[A_i | \bigcap_{j\in S} \overline{A_j}] = \frac{\Pr[A_i \cap \bigcap_{j\in S} \overline{A_j}]}{\Pr[\bigcap_{j\in S} \overline{A_j}]} &\leq \frac{\Pr[A_i \cap \bigcap_{j\in S\setminus D_i} \overline{A_j}]}{\Pr[\bigcap_{j\in S\setminus D_i} \overline{A_j}] \cdot \Pr[\bigcap_{k\in S\cap D_i} \overline{A_k} | \bigcap_{j\in S\setminus D_i} \overline{A_j}]} \\ &= \frac{\Pr[A_i]}{\Pr[\bigcap_{k\in S\cap D_i} \overline{A_k} | \bigcap_{j\in S\setminus D_i} \overline{A_j}]},\end{aligned} \tag{6}$$

where the first inequality used that we are taking a product over fewer events, and the second equality used that A_i is mutually independent of events not in D_i. Let $S \cap D_i = \{k_1, \ldots, k_p\}$. Then we can rewrite the probability in the denominator of (6) as

$$\prod_{q=1}^{p} \Pr[\overline{A_{k_q}} | \bigcap_{j\in(S\setminus D_i)\cup\{k_1,\ldots,k_{q-1}\}} \overline{A_j}] \geq \prod_{q=1}^{p}(1 - \gamma_{k_q}) \geq \prod_{j\in D_i}(1 - \gamma_j),$$

[1] It is perhaps more natural to try and prove the result by induction on n rather than on this somewhat weird looking claim. However, if one plays around with this problem, one quickly sees that one needs to prove something like the stated claim.

where the first inequality used the inductive hypothesis and the last step used $k_q \in S \cap D_i \subseteq D_i$ for all q. This together with (6) and the hypothesis $\Pr[A_i] \le \gamma_i \prod_{j\in D_i}(1-\gamma_j)$ implies that $\Pr[A_i|\bigcap_{j\in S}\overline{A_j}] \le \gamma_i$, which is equivalent to saying $\Pr[\overline{A_i}|\bigcap_{j\in S}\overline{A_j}] \ge 1-\gamma_i$. This proves the inductive hypothesis of our claim, and hence proves the result. □

The following version of the local lemma is often sufficient for most applications (and again this is often referred to as "the local lemma").

Corollary 3.2 (Symmetric Lovász Local Lemma). *Let $A_1,\ldots,A_n$ be events and let $D_1,D_2,\ldots,D_n \subseteq [n]$ be such that A_i is mutually independent of $\{A_j : j \notin D_i \cup \{i\}\}$ for all i. If $\Delta \ge 1$ is such that $|D_i| \le \Delta$ and $\Pr[A_i] \le \frac{1}{e(\Delta+1)}$ for all i, then $\Pr[\bigcap \overline{A_i}] > 0$.*

Proof. Observe that for all i we have

$$\frac{1}{\Delta+1}\prod_{j\in D_i}\left(1-\frac{1}{\Delta+1}\right) \ge \frac{1}{\Delta+1}\left(1-\frac{1}{\Delta+1}\right)^{\Delta} \ge \frac{1}{e(\Delta+1)} \ge \Pr[A_i],$$

where the second to last inequality used that $(1-1/x)^{x-1} > 1/e$ for $x \ge 2$. Thus the (asymmetric) local lemma applies with $\gamma_i = \frac{1}{\Delta+1}$ for all i, proving the result. □

We note that this result is essentially best possible. Indeed, consider rolling a fair $(\Delta+1)$-sided dice and let A_i be the event that the dice rolls i. In this case A_i is dependent on all of $D_i = [\Delta+1] \setminus \{i\}$ and we have $\Pr[A_i] = \frac{1}{\Delta+1} > \frac{1}{e(\Delta+1)}$, so the local lemma does not apply (which is good since we have $\Pr[\bigcap \overline{A}_i] = 0$). In particular, this example shows that we can not improve the requirement $\Pr[A_i] \ge \frac{1}{e(\Delta+1)}$ in the symmetric local lemma to $\Pr[A_i] \ge \frac{1}{\Delta+1}$ in general. Thus the hypothesis in the symmetric local lemma is sharp up to a factor of e, and in fact Shearer proved that this factor of e is necessary [134].

3.1 Applications to Ramsey Theory

Our first application of the local lemma will be an asymptotic improvement to our lower bound for Ramsey numbers from Theorem 1.1.

Theorem 3.3 (Spencer [137]). *For all n we have*

$$R(n,n) \ge (1+o(1))\frac{\sqrt{2}n}{e}2^{n/2}.$$

Proof. Uniformly at random color the edges of K_N. For $S \in \binom{[N]}{n}$, let A_S be the event that G contains a monochromatic K_n on S, and as before we note that $\Pr[A_S] = 2^{1-\binom{n}{2}}$. Let D_S consist of all the sets $T \in \binom{[N]}{n} \setminus \{S\}$ such that $|S \cap T| \ge 2$. It is not difficult to see that A_S is mutually independent of $\{A_T : T \notin D_S \cup \{S\}\}$ since the color given to each pair of S is independent of these events. A weak bound gives $|D_S| \le \binom{n}{2}\binom{N}{n-2} - 1 \le n^2(eN/(n-2))^{n-2} - 1$, so by the (symmetric) local lemma we have that $\Pr[\bigcap \overline{A_S}] > 0$ provided $2^{1-\binom{n}{2}} < \frac{1}{en^2}(eN/n-2)^{2-n}$, i.e. if

$$(2en^2)^{1/n-2} \cdot 2^{\binom{n}{2}/n-2} \cdot \frac{n-2}{eN} = (2en^2)^{1/n-2} \cdot 2^{n/2+1/2-1/(n-2)} \cdot \frac{n-2}{eN} < 1,$$

and this happens if $N = (1-\epsilon)\frac{\sqrt{2}n}{e}2^{n/2}$ for any $\epsilon > 0$ provided n is sufficiently large, giving the desired result. $\square$

The local lemma works best if there are few dependencies between events. As such, it performs much better for off-diagonal Ramsey numbers.

Theorem 3.4. *For all n we have*

$$R(3,n) = \Omega(n^2/\log^2 n).$$

Proof. Randomly color each edge of K_N red with probability p and blue otherwise. Given a set $S \in \binom{[N]}{3}$, we let R_S be the event that the vertices of S form a red triangle, and similarly for $T \in \binom{[N]}{n}$ we define B_T. Observe that $\Pr[A_S] = p^3$ and $\Pr[B_T] = (1-p)^{\binom{n}{2}}$.

Given $S \in \binom{[N]}{3} \cup \binom{[N]}{n}$, we define D_S to be the sets of sizes 3 and n which intersect S in at least two vertices. Observe that if $|S| = 3$, then D_S contains at most $3N$ set of size 3 and at most $\binom{N}{n}$ sets of size n, and if $|S| = n$, we have that D_S contains at most $N\binom{n}{2}$ sets of size 3 and at most $\binom{N}{n}$ sets of size n. Our goal now is to choose some parameters γ_S, γ_T so that the (asymmetric) local lemma applies to the R_S, B_T events.

At this point there's a lot of undetermined variables floating around: N, p, γ_S. Let's think about reasonable guesses for how to optimize things. First of all, it seems clear that we probably want two parameters γ_3, γ_n such that we set $\gamma_S = \gamma_{|S|}$ when applying the local lemma. With this we in particular need

$$p^3 \le \gamma_3(1-\gamma_3)^{3N}(1-\gamma_n)^{\binom{N}{n}}. \tag{7}$$

In particular we need $\gamma_3 \ge p^3$, so let's naively take $\gamma_3 = Cp^3$ for some large constant C. Given this, we also need $\gamma_n \le c\binom{N}{n}^{-1}$ in order to have the $(1-\gamma_n)^{\binom{N}{n}}$ term be no larger than a constant. If we take $\gamma_n = c\binom{N}{n}^{-1}$, we see that (7) is satisfied provided $p = o(N^{-1/3})$ and c, C are chosen appropriately.

The other condition we need to satisfy is

$$(1-p)^{\binom{n}{2}} \le \gamma_n(1-\gamma_3)^{N\binom{n}{2}}(1-\gamma_n)^{\binom{N}{n}},$$

and by plugging in our choices for γ_3, γ_n and the assumption that p must be fairly small, we essentially need to have

$$e^{-p\binom{n}{2}} \le (n/N)^n \cdot e^{-p^3 N\binom{n}{2}},$$

and for this to hold we in particular need something like $p\binom{n}{2} \ge p^3 N\binom{n}{2}$, i.e. $p = O(N^{-1/2})$. Taking $p = c'N^{-1/2}$, we see that we also need roughly

$$p\binom{n}{2} \approx c'N^{-1/2}n^2 \ge n\log(N/n).$$

Assuming $N \ge n^{1+\epsilon}$ for some small $\epsilon > 0$, this reduces to $N^{1/2} \le n/\log n$, i.e. $N = n^2/(\log n)^2$.

Thus in total, a heuristic argument suggests that we can apply the local lemma with $N = \Theta(n^2/(\log n)^2)$ by taking $p = \Theta(N^{-1/2})$, $\gamma_3 = \Theta(N^{-3/2})$, and $\gamma_n = \Theta(\binom{N}{n})$. And indeed, a careful analysis shows that this will work out for n sufficiently large. $\square$

We note that the bound of $n^2/\log^2 n$ is the best one can do using this approach. However, it turns out that $R(3,n) = \Theta(n^2/\log n)$. This improved lower bound was originally proved by Kim [101]. The idea of their proof was to start with a K_N which is entirely colored blue, and then to iteratively randomly pick an edge of K_N and color it red if it does not create a red triangle. A careful analysis shows that with positive probability the final graph at the end contains no large blue clique, and it contains no red clique by construction. We will see a shorter proof of this lower bound in **a later section (Jacques-Dhruv spectral expanders) if I ever write this up**.

We close with a problem from "generalized Ramsey theory." To this end, we define a (p,q)-*coloring* of a graph G to be an edge-coloring of G such that every p-clique of G receives at least q distinct colors. For example, a $(p,2)$-coloring is just an edge-coloring without any monochromatic K_p's. We define the generalized Ramsey number $GR(n,p,q)$ to be the smallest number of colors needed in a (p,q)-coloring of K_n. Until very recently, the best known bounds for $GR(n,p,q)$ come from an old result of Erdős and Gyárfás [58] proved using the local lemma.

Theorem 3.5 ([58])**.** *For all p,q with $p \geq 3$ and $1 \leq q \leq \binom{p}{2}$, we have*

$$GR(n,p,q) \leq p^{\frac{p^2}{\binom{p}{2}-q+1}} n^{\frac{p-2}{\binom{p}{2}-q+1}}.$$

We note that the original result of [58] only stated the bound $GR(n,p,q) = O_{p,q}(n^{(p-2)/(\binom{p}{2}-q+1)})$ with no explicit dependencies on p,q for the implicit constants. Here we emphasize this dependency only because we will later improve upon it using some more advanced machinery; see our forthcoming Theorem 21.5 for more.

Proof. We wish to show that there exists some (p,q)-coloring of K_n using at most $C := p^{\frac{p^2}{\binom{p}{2}-q+1}} n^{\frac{p-2}{\binom{p}{2}-q+1}}$ edges, and for this we consider a uniform random coloring of the edges of K_n using the colors $\{1,2,\ldots,C\}$.

Given a set P of p vertices, we let A_p denote the event that our random coloring gives less than q distinct colors to the edges between vertices of P. Observe that

$$\Pr[A_P] \leq C^{q-1}(q-1)^{\binom{p}{2}} \cdot C^{-\binom{p}{2}},$$

since out of the $C^{\binom{p}{2}}$ equally likely coloring of the edges of P, at most $C^{q-1}(q-1)^{\binom{p}{2}}$ use at most $q-1$ colors (since such a coloring can be identified by first choosing the set of colors used in at most C^{q-1} ways, and then these colors can be assigned to the edges in at most $(q-1)^{\binom{p}{2}}$ ways).

It is not difficult to see that each A_P event is independent of all but at most $\binom{p}{2}\binom{n}{p-2} - 1$ other such events, so by the symmetric local lemma we have $\Pr[\bigcap \overline{A_P}] > 0$ provided

$$e\binom{p}{2}\binom{n}{p-2} \cdot C^{q-1}(q-1)^{\binom{p}{2}} \cdot C^{-\binom{p}{2}} \leq 1.$$

Using $e\binom{p}{2}\binom{n}{p-2} \leq p^2 n^{p-2}$ for all $p \geq 3$ and $(q-1)^{\binom{p}{2}} \leq p^{p^2-p} \leq p^{p^2-2}$, we see that this bound does indeed hold for $C = p^{\frac{p^2}{\binom{p}{2}-q+1}} n^{\frac{p-2}{\binom{p}{2}-q+1}}$, so we conclude that there exists some C-coloring where none of the A_P events occur, giving the desired (p,q)-coloring using at most C colors. $\square$

3.2 Related Lemmas

While the local lemma is very powerful, there are certain circumstances where it doesn't give you quite what you want. Fortunately there are many other lemmas which allow one to prove bounds on $\Pr[\bigcap \overline{A_i}]$ even when the A_i depend on each other in some way. For example, it is not too difficult to generalize the local lemma as follows (and as an exercise the reader should convince themselves that they can prove this result).

Theorem 3.6. *Let $A_1, \ldots, A_n$ be events. Assume there exists partitions $D_i \cup E_i = [n] \setminus \{i\}$ for all i and real numbers $0 \le \delta, \gamma \le 1$ such that $\gamma(1-\gamma)^{|D_i|} \ge \delta$ and for all $E \subseteq E_i$ we have $\Pr[A_i | \bigcap_{e \in E} \overline{A}_e] \le \delta$ and . Then*

$$\Pr[\bigcap \overline{A}_i] \ge (1-\gamma)^n > 0.$$

Note that when $\delta = \gamma$ we more or less recover Theorem 3.1 when $\gamma_i = \gamma$ for all i. The power here is that we allow each A_i to possible be dependent of every event, but it is not "very dependent" on the events of E_i.

Another result in a similar spirit as the local lemma is Janson's inequality. Given a set $S \subseteq X$ and $\vec{p}$, let $\mathcal{A}_S$ be the set containing

Theorem 3.7 (Janson's inequality). *Let H be a hypergraph on a set V, and let V_p be the set obtained by including each vertex of V independently and with probability p. Let A_i denote the event that V_p contains the ith edge of H and define*

$$\mu = \sum \Pr[A_i], \ \ \Delta = \sum_{(S_i, S_j): S_i \cap S_j \neq \emptyset} \Pr[A_i \cap A_j].$$

Then

$$\prod_i \Pr[\overline{A_i}] \le \Pr[\bigcap_i \overline{A_i}] \le e^{-\mu + \frac{\Delta}{2}}.$$

Note that if all of the edges of H are disjoint, then these bounds are roughly $e^{-\mu} \le \Pr[\bigcap_i \overline{A_i}] \le e^{-\mu/2}$. Again there are many variants of Theorem 3.7 which are useful in different situations.

4 Concentration Inequalities

Up to this point, we have largely applied the probabilistic method by showing that random variables X are sufficiently large with some arbitrarily small positive probability. In more advanced uses of the method, one often needs to go beyond this and say that X is in fact fairly likely to be quite close to its expectation. There are a plethora of tools throughout the probability literature for accomplishing exactly this goal, and below we survey some of the most common ones used in probabilistic and extremal combinatorics. We will omit the proofs of all but the simplest of these results, focusing instead on their main applications. We refer the reader to the book of Dubhashi and Panconesi [52] for complete proofs, and we note that the appendix of [52] consists of a very nice summary of these inequalities as well as many of their generalizations.

4.1 Markov and Chebyshev

Perhaps the most famous (one-sided) concentration inequality is Markov's inequality. We already saw this around the proof of Theorem 2.2, but for good measure we'll formally state it here.

Lemma 4.1 (Markov's inequality)**.** *If X is a non-negative real-valued random variable, then for all $\lambda > 0$ we have*

$$\Pr[X \geq \lambda] \leq \frac{\mathbb{E}[X]}{\lambda}.$$

In particular, if X is integer-valued, then

$$\Pr[X \neq 0] \leq \mathbb{E}[X].$$

Proof. For simplicity we only prove the result when X is integer valued. In this case we have

$$\Pr[X \geq \lambda] = \sum_{k \geq \lambda} \Pr[X = k] \leq \sum_{k \geq \lambda} \frac{k}{\lambda} \cdot \Pr[X = k] = \frac{\mathbb{E}[X]}{\lambda}.$$

The second statement follows by taking $\lambda = 1$. □

The "in particular" part of this lemma is probably the most common usage of Markov's inequality. To reiterate, this says that $\mathbb{E}[X] \to 0$ implies $X = 0$ with high probability, and this application of Markov's inequality is often known as *the first moment method*.

Unfortunately it is not true in general that $\mathbb{E}[X] \to \infty$ implies $X > 0$ with high probability (e.g. take $X = n$ with probability $n^{-1/2}$ and $X = 0$ otherwise). However, for many reasonable examples this implication does hold. Often one can show this by utilizing Chebyshev's inequality. We recall that the variance of a random variable is $\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$.

Lemma 4.2 (Chebyshev's inequality)**.** *Let X be a real-valued random variable with $\mathrm{Var}(X) = \sigma^2$. Then for all $\lambda > 0$, we have*

$$\Pr[|X - \mathbb{E}[X]| \geq \lambda\sigma] \leq \frac{1}{\lambda^2}.$$

Proof. We have

$$\Pr[|X - \mathbb{E}[X]| \geq \lambda\sigma] = \Pr[(X - \mathbb{E}[X])^2 \geq \lambda^2\sigma^2] \leq \frac{1}{\lambda^2},$$

where this last step used Markov's inequality applied to the (non-negative) random variable $Y := (X - \mathbb{E}[X])^2$ after noting that $\mathbb{E}[Y] = \sigma^2$ by definition. □

Morally speaking, Chebyshev's inequality says that if $\sigma = o(\mathbb{E}[|X|])$, then X is close to its expectation with high probability. The usage of Chebyshev's inequality is often referred to as *the second moment method*.

Chebyshev's inequality is perhaps the most flexible concentration inequality out there, in that it applies to arbitrary random variables with finite second moments. While the second moment method can often be used to give good enough bounds for applications (see for example our forthcoming proof sketch of Theorem 20.1), the extreme flexibility of the method means that the bounds obtained from it will often be far from the true concentration behavior of our random variables. If one needs better bounds than what Chebyshev gives, then one needs to employ a concentration inequality which applies to a narrow set of random variables. We consider three such results in the following subsections.

4.2 The Chernoff Bound

The Chernoff bound says that binomial random variables have exponential concentration around their means. This version is incorrect, need to adjust.

Theorem 4.3. *Let $X_1, \ldots, X_n$ be independent Bernoulli random variables each with probability of success p, and let $X = \sum X_i$. Then for all $\lambda > 0$,*

$$\Pr[|X - pn| \geq \lambda pn] < 2e^{-\lambda^2 pn/2}.$$

Sketch of Proof. Observe that for all $\lambda, t > 0$, we have

$$\Pr[X \geq (1+\lambda)pn] = \Pr[e^{tX} \geq e^{t(1+\lambda)pn}] \leq \mathbb{E}[e^{tX}]e^{-t(1+\lambda)pn},$$

with this last step using Markov's inequality. We note that $e^{tX} = \sum \frac{t^m \mathbb{E}[X^m]}{m!}$ is the *moment generating function* of X, and it is a common trick in probability to rephrase inequalities in terms of e^{tX}. And indeed, because the X_i are all independent, we have

$$\mathbb{E}[e^{tX}] = \prod \mathbb{E}[e^{tX_i}] = (e^t p + (1-p))^n.$$

Thus we are left with the problem of choosing t so that $\frac{e^t p + (1-p)}{e^{-t(1+\lambda)p}}$ is minimized. One can do this using calculus, and this will give $\Pr[X \geq (1+\lambda)pn] < e^{-\lambda^2 pn/2}$. The same argument gives $\Pr[X \leq (1-\lambda)pn] < e^{-\lambda^2 pn/2}$, and combining these inequalities gives the desired result. □

The Chernoff bound can be generalized, for example, by replacing the bernoulli random variables with any bounded random variable, see [52].

Many random variables in probabilistic combinatorics end up being binomial random variables, such as the number of edges in $G_{n,p}$, and as such the Chernoff bound is heavily used in practice. Here we look at a quick application of this result to a problem in discrepency theory.

Given a hypergraph H and partition $V(H) = R \sqcup B$, we define the *discrepancy* of the partition by

$$\operatorname{disc}(H, R, B) = \max_{e \in E(H)} \left| |e \cap R| - |e \cap B| \right|,$$

and define the discrepancy of the hypergraph by $\operatorname{disc}(H) = \min_{R,B} \operatorname{disc}(H, R, B)$. In other words, $\operatorname{disc}(H)$ measures how well one can partition the vertex set so that each edge has about the same number of vertices from each part.

Theorem 4.4. *If H is an r-uniform hypergraph with m edges, then* $\operatorname{disc}(H) \le 2\sqrt{r \log(2m)}$.

If H is a clique on $2r-1$ vertices, then $\operatorname{disc}(H) = r$ and $m \approx 4^r$, so this result is essentially best possible for general H.

Proof. Assign each vertex of H to R or B independently and with probability $\frac{1}{2}$. For $e \in E(H)$, let A_e be the event that

$$\left| |e \cap R| - \frac{1}{2} r \right| \ge \sqrt{r \log(2m)} = 2\sqrt{\frac{\log(2m)}{r}} \cdot \frac{1}{2} r.$$

Because $|e \cap R|$ has a binomial distribution, the Chernoff bound gives $\Pr[A_e] < 2e^{-\log(2m)} = m^{-1}$, and by a union bound we have $\Pr[\bigcup A_e] < 1$. Thus with positive probability, there exists a partition R, B such that none of the A_e occur. This means $\operatorname{disc}(H, R, B) \le 2\sqrt{r \log 2m}$, proving the result. $\square$

Much more can be said about discrepancy problems, see [8, Chapter 13].

4.3 Martingales

We say that a sequence of real-valued random variables $X_0, X_1, \ldots$ is a *martingale* if $\mathbb{E}[X_{i+1}|X_i] = X_i$ for all i. One important class of Martingales, called Doob martingales, are defined as follows. Given random variables $Y_1, \ldots, Y_m$ and a real-valued function f, let

$$X_i = \mathbb{E}[f(Y_1, \ldots, Y_m)|Y_1, \ldots, Y_i].$$

It is not too difficult to show that any sequence of random variables X_i defined in this way is indeed a martingale.

One of the most common classes of (Doob) martingales in probabilistic combinatorics are the edge-exposure martingales. In this case, Y_i denotes the indicator random variable which is 1 if the ith pair of vertices in $G_{n,p}$ is an edge (where the pairs are ordered in some arbitrary way). Intuitively in this situation we think of revealing the edges of $G_{n,p}$ one at a time, and X_i denotes the value that we expect f to be after we reveal all of the remaining edges.

Let us look at the very concrete case of the edge-exposure martingale when $n = 3$ and f is the number of triangles in $G_{3,p}$. With this we have

$$X_0 = \mathbb{E}[f] = p^3, \ X_1 = \mathbb{E}[f|Y_1] = p^2 Y_1,$$

$$X_2 = \mathbb{E}[f|Y_1, Y_2] = pY_1Y_2, \ X_3 = \mathbb{E}[f|Y_1, Y_2, Y_3] = Y_1Y_2Y_3.$$

The main concentration result for martingales is Hoeffding's inequality.

Theorem 4.5 (Hoeffdings's inequality). *Let $X_0, \ldots$ be a martingale with $|X_i - X_{i-1}| \le \alpha_i$ for all i. Then for all $\lambda > 0$, we have*

$$\Pr[|X_m - X_0| \ge \lambda] < 2e^{\frac{-2\lambda^2}{\sum \alpha_i^2}}.$$

Sketch of Proof. Let $Y_i = X_i - X_{i-1}$. Similar to the proof of the Chernoff bound, we have

$$\Pr[X_m - X_0 \ge \lambda] = \Pr[e^{t(X_m - X_0)} \ge e^{t\lambda}] \le \mathbb{E}[e^{t\sum_{i=1}^m Y_i}]e^{-t\lambda}.$$

We claim that this expectation is at most $e^{\frac{1}{8}t^2\sum_{i=1}^m \alpha_i^2}$. Indeed, we can use conditional expectations to write

$$\mathbb{E}[e^{t\sum_{i=1}^m Y_i}] = \mathbb{E}\left[\mathbb{E}[e^{t\sum_{i=1}^m Y_i}|X_0, \ldots, X_{m-1}]\right] = \mathbb{E}[e^{t\sum_{i=1}^{m-1} Y_i} \cdot \mathbb{E}\left[e^{tY_m}|X_0, \ldots, X_{m-1}\right]],$$

where this last step used that Y_i with $i < m$ is fixed given $X_0, \ldots, X_{m-1}$. Observe that conditional on $X_0, \ldots, X_{m-1}$, we have $\mathbb{E}[Y_m] = 0$ (due to the martingale property) and $|Y_m| \le \alpha_m$ (due to the hypothesis of the theorem). One can show that for random variables of this form, the expected value of its moment generating function is at most $e^{\alpha_m^2 t^2/8}$. One gets the claim by repeating this argument inductively on the remaining terms.

In total, we have for any $t > 0$ that

$$\Pr[X_m - X_0 > \lambda\sqrt{m}] \le e^{\frac{1}{8}t^2\sum \alpha_i^2 - t\lambda}.$$

Taking $t = 4\lambda/\sum \alpha_i^2$ gives $\Pr[X_m - X_0 \ge \lambda] < e^{\frac{-2\lambda^2}{\sum \alpha_i^2}}$. A symmetric argument shows $\Pr[X_m - X_0 \le \lambda] < e^{\frac{-2\lambda^2}{\sum \alpha_i^2}}$ (this can also be seen by considering the martingale $X_i' := -X_i$ and applying the first inequality), which gives the result. ☐

In the special case where $\alpha_i = 1$ for all i, this result is referred to as Azuma's inequality[1].

Corollary 4.6 (Azuma's inequality). *Let $X_0, \ldots$ be a martingale which satisfies $|X_i - X_{i-1}| \le 1$ for all i. Then for all $\lambda > 0$, we have*

$$\Pr[|X_m - X_0| \ge \lambda\sqrt{m}] < 2e^{-2\lambda^2}.$$

[1]The naming convention for these inequalities are all over the place: some people call these Hoeffding's inequalities, others Azuma (which is probably the most popular name in the combinaotrics community), some Azuma-Hoeffding, and yet others Hoeffding-Azuma.

There are many generalizations of the Hoeffding's inequality which weakens the hypothesis that $|X_{i+1} - X_i| \le \alpha_i$. For example, it suffices to have that this difference holds in expectation, or that it holds with high probability. Again, see [52] for details.

One application of Azuma's inequality is the following.

Proposition 4.7. *We have*

$$\Pr[|\chi(G_{n,p}) - \mathbb{E}[\chi(G_{n,p})| \ge \lambda\sqrt{n}] < 2e^{-\lambda^2/2}.$$

We note that this result tells us that $\chi(G_{n,p})$ is concentrated around its expectation, but it gives no indication of what this expectation is. This is a common phenomenon when applying concentration inequalities.

Proof. We consider a vertex-exposure martingale, i.e. a Doob martingale $f(Y_1, \ldots, Y_n)$ where Y_i is the set of vertices $j > i$ in $G_{n,p}$ which are adjacent to i. In particular, taking $f = \chi$ and $X_i = \mathbb{E}[f|Y_1, \ldots, Y_i]$ gives $X_0 = \mathbb{E}[\chi(G_{n,p})]$ and $X_n = \chi(G_{n,p})$. It is clear that each time we reveal a set Y_i that the expected chromatic number changes by at most 1, i.e. $|X_{i+1} - X_i| \le 1$ for all i. Thus Azuma's inequality applies, giving the result.

$\square$

We note that one could try and prove this result using an edge-exposure martingale instead of a vertex-exposure martigale, but this approach gives essentially trivial bounds. In general, when using martingales you want to reveal information in as few rounds as possible, while also making it so that the information you reveal can't dramatically change your function each round.

While Proposition 4.7 says nothing about $\mathbb{E}[\chi(G_{n,p})]$, it is well known that this value is asymptotic to $\frac{n}{2\log_{1/(1-p)} n}$ for any fixed p. This was first proven by Bollobás using a clever martingale argument. Much more can be said about $\chi(G_{n,p})$, see for example the paper by Heckel and Riordan [88] which, in addition to surveying many of the known results on $\chi(G_{n,p})$, shows that the concentration in Proposition 4.7 is in some sense close to best possible.

There are a number of variants of all of the concentration inequalities stated in this chapter. One particular version of Azuma that we will need at some point is the following.

Lemma 4.8 ([107])**.** *Let $X_0, \ldots$ be a martingale which satisfies $|X_i - X_{i-1}| \le \alpha$ for all i. Then for all $\delta \in [0, 1]$, we have*

$$\Pr[|X_m - X_0| \ge \delta\alpha m] < e^{-\delta^2\alpha m/6c}.$$

Note that $|X_m - X_0| \le \alpha m$ deterministically, so $\delta\alpha m$ is at least a δ fraction of the mean of $X_m - X_0$, and as such this is referred to as the "multiplicative Azuma inequality" (since its error term is multiplicative relative to the expectation as opposed to additive).

4.4 Talagrand's Inequality

Let $\Omega = \Omega_1 \times \cdots \times \Omega_n$ be a product of probability spaces. For $\alpha = (\alpha_1, \ldots, \alpha_n)$ a vector of non-negative real numbers, we define the *weighted Hamming distance* d_α on Ω by $d_\alpha(x, y) = \sum_{i:x_i \ne y_i} \alpha_i$. For example, $\alpha = (1, \ldots, 1)$ gives the usual Hamming distance on product spaces.

Given α as above, a set $A \subseteq \Omega$, and a non-negative number t, we define

$$A_{\alpha,t} = \{x : \exists y \in A,\ d_\alpha(x,y) \le t\}.$$

That is, $A_{\alpha,t}$ is the set of points in Ω which are within distance t of A. We also define $\overline{A} = \Omega \setminus A$. Our goal is to prove "isoperimetric" inequalities which state that, for any $A \subseteq \Omega$, we have

$$\Pr[A] \cdot \Pr[\overline{A_{\alpha,t}}] \le f(t),$$

where f is some rapidly shrinking function. Isoperimetric inequalities are intimately related to concentration inequalities. For example, a corollary of an inequality as above is that if $\Pr[A] \ge \frac{1}{2}$, then $\Pr[\overline{A_{\alpha,t}}] \le 2f(t)$ (i.e., most of Ω is concentrated around A). On the other hand, one can prove isoperimetric inequalities by using concentration inequalities.

Proposition 4.9. *For Ω a product space, $A \subseteq \Omega$, and α such that $\sum \alpha_i^2 = 1$, we have for all t that*

$$\Pr[A] \Pr[\overline{A_{\alpha,t}}] \le 4e^{-t^2}$$

.

Proof. Define the function $f : \Omega \to \mathbb{R}$ by $f(y) = d_\alpha(y, A)$. Let $Y = (Y_1, \ldots, Y_n)$ be chosen according to the probability distribution on Ω and let $X_i = \mathbb{E}[f(Y)|Y_1, \ldots, Y_i]$. Observe that $X_n = 0$ iff $Y \in A$ and $X_n > t$ iff $Y \in \overline{A_{\alpha,t}}$, and also that $|X_i - X_{i-1}| \le \alpha_i$. Thus Hoeffding's inequality implies

$$\begin{aligned}
\Pr[A] \Pr[\overline{A_{\alpha,t}}] &= \Pr[X_m = 0] \Pr[X_m > t] \\
&\le \Pr[|X_m - X_0| \ge X_0] \Pr[|X_m - X_0| > t - X_0] \\
&\le 4e^{-2X_0^2 - 2(t-X_0)^2} \le 4e^{-t^2},
\end{aligned}$$

where this last step used that the exponent is maximized when $X_0 = \frac{1}{2}t$. □

A remarkable result of Talagrand shows that Proposition 4.9 essentially holds even when comparing A with the set of points which are at least distance t from A for *some* choice of α.

Theorem 4.10 (Talagrand's inequality)**.** *For all $A \subseteq \Omega$ and $t \ge 0$, we have*

$$\Pr[A] \Pr\left[\overline{\bigcap_\alpha A_{\alpha,t}}\right] \le e^{-t^2/4},$$

where the intersection ranges over all α with $\sum \alpha_i^2 = 1$.

Again we emphasize that $\overline{\bigcap_\alpha A_{\alpha,t}}$ can be much larger than $\overline{A_{\alpha,t}}$ for any given α, but still essentially the same bound as in Proposition 4.9 holds. We omit the proof of Theorem 4.10, and we refer the reader to [8] for a direct proof, and to [52] for a longer, but perhaps more enlightening argument.

We note that Talagrand's inequality is often stated in the following equivalent form: Given $x \in \Omega$ and $A \subseteq \Omega$, define $d'(x, A) = \min_{y \in A} \max_\alpha d_\alpha(x, y)$, where the maximum ranges over all

α with $\sum \alpha_i^2 = 1$. Note that having $d'(x,A) \le t$ is equivalent to saying that for all α there exist $y \in A$ with $d_\alpha(x,y) \le t$, which is equivalent to saying $x \in \bigcap_\alpha A_{\alpha,t}$. Thus Theorem 4.10 can be seen as an isoperemetric inequality with respect to the pseudo-distance d'.

Talagrand's inequality has a number of applications to concentration of random variables. One particular application is for certifiable functions. For a function $s : \mathbb{R} \to \mathbb{N}$, we say that a real-valued function f defined on a product space Ω is *s-certifiable* if having $f(x) \ge c$ implies that there exists a set $I \subseteq [n]$ of size $s(c)$ such that $f(y) \ge c$ whenever $y_i = x_i$ for all $i \in I$ (that is, the values in position I "certify" that $f(x) \ge c$).

For example, if $f(x) = |\{i : x_i \neq 0\}|$, then f is s-certifiable with $s(c) = c$, since $f(x) \ge c$ implies there exist c coordinates with $x_i \neq 0$, and any y which agrees with x on these coordinates satisfies $f(y) \ge c$. Lastly, we say that a function f is Lipschitz if $|f(x) - f(y)| \le 1$ whenever x, y differ in at most one coordinate.

Corollary 4.11. *If f is an s-certifiable Lipschitz function on the product space Ω and X is chosen according to the probability space Ω, then for all m and $t > 0$ we have*

$$\Pr[f(X) < m - t\sqrt{s(m)}]\Pr[f(X) \ge m] \le e^{-t^2/4}.$$

Proof. Let $A = \{x : f(x) < m - t\sqrt{s(m)}\}$. We claim that $\bigcap_\alpha A_{\alpha,t} \subseteq \{y : f(y) < m\}$.

Assume for contradiction that $y \in \bigcap_\alpha A_{\alpha,t}$ and $f(y) \ge m$. Because f is s-certifiable, there exists a set of $s(m)$ indices I which certifies $f(y) \ge m$. Let $\alpha'_i = \frac{1}{\sqrt{s(m)}}$ if $i \in I$ and $\alpha'_i = 0$ otherwise. Because $y \in \bigcap_\alpha A_{\alpha,t} \subseteq A_{\alpha',t}$, there exists some $x \in A$ such that $d_{\alpha'}(x,y) \le t$, i.e. such that restricted to I, the vectors x, y differ in at most $t\sqrt{s(m)}$ coordinates. Let z be defined by $z_i = y_i$ if $i \in I$ and $z_i = x_i$ otherwise. Then $f(z) \ge m$ by definition of I, and f being Lipschitz implies $f(x) \ge f(z) - t\sqrt{s(m)} \ge m - t\sqrt{s(m)}$, contradicting $x \in A$. This proves the claim.

The contrapositive of the claim implies $\{y : f(y) \ge m\} \subseteq \overline{\bigcap_\alpha A_{\alpha,t}}$, so the first result follows from Talagrand's inequality. $\square$

We emphasize that Proposition 4.9 is too weak to prove Corollary 4.11: we genuinely have to make use of the fact that Talagrand's inequality allows us to choose a different distance function for each choice of y.

In most applications, one applies Corollary 4.11 where either m is a median, i.e. $\Pr[f(X) \ge m] = \frac{1}{2}$, or where $m - t\sqrt{s(m)}$ is a median. While medians are hard to estimate directly, a concentration result like that of Corollary 4.11 can usually be used to show that the median and expectation must be close to each other, see for example [52, Problem 11.4].

As an application, let $X = (X_1, \ldots, X_n)$ be a random vector with each X_i distributed uniformly on $[0,1]$. Let $f(X)$ denote the length of a longest increasing subsequence, i.e. the largest k such that there exist indices with $X_{i_1} < X_{i_2} < \cdots < X_{i_k}$. Note that f is Lipschitz and is s-certifiable with $s(c) = c$, so if m is a median we conclude $\Pr[f(X) < m - t\sqrt{m}] \le 2e^{-t^2/4}$. It is well known that $\mathbb{E}[f(X)] \sim 2\sqrt{n}$, so at least heuristically, this argument suggests $f(X)$ is highly concentrated around $2\sqrt{n} + \Theta(n^{1/4})$ (and it's not hard to make this more precise). In contrast, if one attempted to get concentration results for $f(X)$ by utilizing martingales, one would conclude that $f(X)$ is highly concentrated around $\Theta(n^{1/2})$, which is significantly weaker. One can literally dedicate an entire book to the longest increasing subsequence problem, see Romik [131] for more on this topic.

Part II

Further Probabilistic Methods

5 Dependent Random Choice

The following is all based off of the excellent survey by Fox and Sudakov [63]. Throughout this section we denote the common neighborhood of a set of vertices S by $N(S)$, i.e. $N(S) = \{u : u \in N(v)\ \forall v \in S\}$. Before we explain what dependent random choice is, let's first see an example of it in action.

Lemma 5.1. *Let G be an n-vertex graph with average degree at least d. For any choice of integers m, r, t, there exists a set $U \subseteq V(G)$ such that every r-subset of U has at least m common neighbors, and such that*

$$|U| \geq \frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t.$$

Proof. The statement of the result suggests how we should prove it: we'll randomly pick a set W which will have expected size at least d^t/n^{t-1}, and then we'll use the method of alterations to delete from W a set of a bad vertices, which in expectation will have size at most $\binom{n}{r}(m/n)^t$. The key twist is that we don't start by, say, defining W to include each vertex independently and with probability $p = d^t/n^t$, but instead W will end up depending on a different random set T.

To this end, let T be the random set obtained by uniformly at random selecting t vertices with repetition (i.e. each vertex is equally likely to be the ith vertex added to T, and in total T has size at most t), and define $W = N(T)$. The probability that a given vertex v is included in W is exactly $(d(v)/n)^t$, so by linearity of expectation and convexity we find that

$$\mathbb{E}[|W|] = \sum (d(v)/n)^t \geq d^t/n^{t-1}.$$

We say that a set of vertices $S \subseteq V(G)$ of size r is bad if $|N(S)| \leq m$. The probability that W contains a given bad set S is at most $(m/n)^t$ (since $S \subseteq W$ iff $T \subseteq N(S)$). Thus the expected number of bad sets of W is at most $\binom{n}{r}(m/n)^t$. If we let U be the set obtained by deleting a vertex from each bad set of W, then it has the desired properties by construction and in expectation it has the desired size, so such a choice of U exists. $\square$

Again, the key idea of this proof is that instead of defining W by including each vertex independently and with probability $p = d^t/n^t$, we instead formed it so that, on average, each vertex has probability at least p of being added, but the vertices are added in a very dependent way. In particular, the dependent way that W was generated made it more likely to have our desired property (i.e., we generated W by taking a common neighborhood, which made it less likely for W to contain sets of vertices with small common neighborhoods).

We can use Lemma 5.1 to prove some bounds on Turán numbers by using the following embedding lemma.

Lemma 5.2. *Let F be a bipartite graph on $A \cup B$ with $|A| = a,\ |B| = b$ such that the vertices in B all have degree at most r. If G is a graph which contains a set U such that $|U| = a$ and such that any subset of U of size r contains at least $a + b$ common neighbors, then G contains F as a subgraph.*

Proof. We define an injective homomorphism ϕ from $V(F)$ to $V(G)$ as follows. Choose $\phi|_A$ to be an arbitrary bijection onto U. For each $v \in B$ that has yet to be assigned, choose $\phi(v)$ to be any common neighbor of $\phi(N_F(v))$ which has yet to be assigned by ϕ. Note that there exist at least $a+b$ common neighbors of $\phi(N_F(v))$, so there certainly exists one which has yet to be assigned. This mapping gives the result. $\square$

With this we can quickly prove the following.

Theorem 5.3 (Füredi [72]; Alon, Krivelevich, Sudakov [6])**.** *If F is a bipartite graph on $A \cup B$ such that the vertices of B all have degree at most r, then*

$$\mathrm{ex}(n, F) < 3(a+b)n^{2-1/r}.$$

Observe that this result generalizes Kővári-Sós-Turán, at least in terms of order of magnitude.

Proof. Assume G is an n-vertex F-free graph with average degree $d = 6(a+b)n^{1-1/r}$. By Lemma 5.2, we would be done if we could find a set U of size at least a such that every subset of size r had at least $m = a+b$ common neighbors. By Lemma 5.1, for any t we can find a set U with these properties of size at least

$$\frac{d^t}{n^{t-1}} - \binom{n}{r}\left(\frac{m}{n}\right)^t \geq (6a+6b)^t n^{1-t/r} - (e/r)^r (a+b)^t n^{r-t}.$$

We see that taking $t = r$ makes the powers of n on both sides equal, and in total this gives a set of size at least

$$(6a+6b)^r - (e(a+b)/r)^r.$$

Note that $6(a+b) \geq \frac{1}{2}(e(a+b)/r)$, so this is at least $\frac{1}{2}(6a+6b)^r \geq a$. We have thus found our desired set U, which together with Lemma 5.2 gives a copy of F in G, a contradiction. $\square$

Another application of this method is to subdivisions. We define the 1-subdivision H^* of a graph H to be the graph obtained by replacing each edge of H by a P_2 (i.e. by inserting a new vertex in the middle of each edge). Note that subdivisions are bipartite graphs with all of its $e(H)$ new vertices having degree 2. Thus the previous theorem gives $\mathrm{ex}(n, K_a^*) = O(a^2 n^{3/2})$. It turns out that one can significantly improve upon this dependency of a.

Theorem 5.4 (Alon, Krivelevich, Sudakov [6])**.** *For all a we have*

$$\mathrm{ex}(n, K_a^*) = O(an^{3/2}).$$

Note that this only gives a reasonable bound when $a = O(n^{1/2})$, which makes sense since K_a^* has about a^2 vertices and thus can always be avoided by an n-vertex graph if $a \gg n^{1/2}$.

Unfortunately Lemma 5.1 on its own is not enough to prove Theorem 5.3, essentially because the size of U that we're guaranteed is too small. We can increase the size of U by demanding slightly weaker conditions for it to have, i.e. we only need that most pairs have many common neighbors[1]. More precisely, we use the following.

[1] This is a common situation that happens in applications of dependent random choice, though the exact way you weaken the conditions of Lemma 5.1 depends on the particular problem at hand.

Lemma 5.5. *Let G be an n-vertex graph with $an^{3/2}$ edges. Then G contains a subset of vertices U with $|U| = a$ such that for all $1 \le i \le \binom{a}{2}$, there are less than i pairs of vertices in U with fewer than i common neighbors in $V(G) \setminus U$.*

For example, this says that every pair of vertices of U has at least one common neighbor outside of U, and that there is at least one pair which has at least a common neighbors outside of U.

Add more intuition for the proof.

Proof. For simplicity we assume n is even, and by losing at most half of our edges we can assume that G is bipartite on $V_1 \cup V_2$ with $|V_1| = |V_2| = n/2$. Without loss of generality we can assume $\sum_{v \in V_1} d(v)^2 \le \sum_{v \in V_2} \sum d(v)^2$.

Let T be a random set obtained by including two vertices uniformly at random from V_1 with replacement. Let $W = N(T)$ and $X = |W|$. Similar to our computation before, we find

$$\mathbb{E}[X] = \sum_{v \in V_2} (d(v)/(n/2))^2 \ge 4n^{-2} \cdot (n/2)(an^{1/2})^2 = 2a^2.$$

Given distinct vertices $x, y \in V_2$, we define $f(x,y) = \frac{1}{|N_{V_1}(x,y)|}$ and we let $Y = \sum_{x,y \in W} f(x,y)$. Observe that

$$\begin{aligned}\mathbb{E}[Y] &= \sum_{x,y \in V_2} f(x,y) \cdot \Pr[x,y \in W] = \sum_{x,y \in V_2} \frac{1}{|N_{V_1}(x,y)|} \cdot \left(\frac{|N_{V_1}(x,y)|}{n/2}\right)^2 = 4n^{-2} \sum_{x,y \in V_2} |N_{V_1}(x,y)| \\ &= 4n^{-2} \sum_{z \in V_1} \binom{d(z)}{2} \le 2n^{-2} \sum_{z \in V_1} d(z)^2 \le 2n^{-2} \sum_{z \in V_2} d(z)^2 = \frac{1}{2}\mathbb{E}[X].\end{aligned}$$

With this we see $\mathbb{E}[X - \mathbb{E}[X]/2 - Y] \ge 0$, and thus there exists a choice of T such that $X \ge Y$ and $X \ge \mathbb{E}[X]/2 \ge a^2$.

The trick now is to take $U \subseteq W$ a set of size exactly a uniformly at random, and let $Y' = \sum_{x,y \in U} f(x,y)$. In this case

$$\mathbb{E}[Y'] = \sum_{x,y \in W} f(x,y) \cdot \Pr[x,y \in U | x,y \in W] \le Y \cdot \frac{a(a-1)}{X(X-1)} \le X \cdot (a/X)^2 \le 1.$$

Thus there exists a choice of U such that $Y' \le 1$. We claim that such a U has the desired properties. Indeed, if there existed i pairs with fewer than i common neighbors, then this would immediately imply $Y' \ge i \cdot \frac{1}{i-1} > 1$, a contradiction. $\square$

Theorem 5.4 follows almost immediately from Lemma 5.5, and we omit its proof.

For our last result, we say that a graph F is *r-degenerate* if every subgraph of F contains a vertex of degree at most r. In this setting we can prove an embedding lemma analogous to Lemma 5.2.

Lemma 5.6. *Let G be a graph with vertex sets U_1, U_2 such that, for $k = 1, 2$, every subset of at most r vertices in U_k contains at least m common neighbors in U_{3-k}. Then G contains every r-degenerate bipartite graph H on m vertices.*

Proof. Let F_1 be an m-vertex r-degenerate bipartite graph on $V_1 \cup V_2$. By definition this means that there exists a vertex $v_1 \in F_1$ such that $d_{F_1}(v_1) \le r$, and that there is some $v_2 \in F_2 := F_1 - v_1$ with $d_{F_2}(v_2) \le r$ and so on. We now define a map $\phi : V_1 \cup V_2 \to U_1 \cup U_2$ with $\phi(V_i) \subseteq U_i$ as follows. Iteratively assume we have defined $\phi(v_m), \phi(v_{m-1}), \ldots, \phi(v_{q+1})$ and that $v_q \in V_i$. Since $S := N(v_q) \cap \{v_m, \ldots, v_{q+1}\}$ has at most r vertices by assumption, the set $\phi(S) \subseteq U_{3-i}$ has at least m common neighbors, so choose $\phi(v_q)$ to be any of these vertices that has yet to be assigned. It is not difficult to see that this gives the desired embedding. □

Motivated by this lemma, we prove the following variant of Lemma 5.1.

Lemma 5.7. *Let $r, m \ge 2$ and let G be an n-vertex graph with at least $mn^{1-1/6r}$ edges. Then G contains two subsets U_1, U_2 such that, for $k = 1, 2$, every subset of r vertices in U_k has at least m common neighbors in U_{3-k}.*

Proof. The rough strategy of the proof is as follows. We will first apply Lemma 5.1 directly to obtain a large set U_1 such that every q-subset of U_1 (with $q > r$) has at least m common neighbors. We then mimic the proof of Lemma 5.1 by choosing a random set $T \subseteq U_1$ of size t and letting $U_2 = N(U_1)$. By choosing an appropriate value of t, the set U_2 will satisfy the condition. Moreover, if $q - t \ge r$, then for any r-subset $S \subseteq U_1$, the set $S \cup T$ has at least m common neighbors, all of which in particular lie in $N(T) = U_2$, so U_1 will also have the desired property.

We now being the formal argument. Apply Lemma 5.1 using $q = 3r$ instead of r, t to get a set U_1 such that every subset of size $3r$ has at least m common neighbors and such that

$$|U_1| \ge \frac{d^{3r}}{n^{3r-1}} - \binom{n}{3r}(m/n)^{3r} \ge m^{3r}n^{1/2} - m^r/(3r)! \ge mn^{1/2}.$$

Now let T be a set obtained by including $t = 2r$ vertices uniformly at random from U_1 with replacement, and let $U_2 = N(T)$. The probability that U_2 contains a set of r vertices which have fewer than m common neighbors in U_1 is at most

$$\binom{n}{r}(m/|U_1|)^{2r} \le \frac{1}{r!} < 1,$$

and in particular there exists a choice of T such that no r-subset of U_2 has fewer than m common neighbors. Note that for any r-subset $S \subseteq U_1$, the set $S \cup T$ has size at most $3r$ vertices, so by construction S has at least m common neighbors which lie in $N(T) = U_2$. Thus U_1, U_2 gives the desired result. □

Combining these two lemmas immediately gives the following.

Theorem 5.8. *If F is an m-vertex r-degenerate graph, then*

$$\mathrm{ex}(n, F) < mn^{2-1/6r}.$$

We note that one can optimize the proof of Lemma 5.7 to improve the exponent of this theorem slightly (by using $(3-2\sqrt{2})r$ instead of $3r$ throughout). However, the end result is still weaker than the best known bound of $\mathrm{ex}(n, F) \leq m^{1/2r} n^{2-1/4r}$ due to Alon, Krivelevich, and Sudakov [6], with their proof more or less being a slight refinement of the argument we gave.

As all of these examples illustrate: if you have a problem that could be magically solved if you had a large set of vertices U such that every r-set of U had many common neighbors, then a variant of dependent random choice might be worth trying out!

6 Coupling

It is often the case that one can understand a random variable X by comparing it to a "similar" random variable Y which is easier to do calculations for. One way to do this to form a *coupling*, i.e. a pair of random variables (X', Y') such that X', Y' have the same distribution as X, Y, respectively, and such that X', Y' have some (nice) relation between them. In particular, we will consider two beautiful couplings around random graphs, one relating $G_{n,p}$ to random digraphs, and the other to random hypergraphs.

6.1 Graphs and Digraphs

Define the random digraph $D_{n,p}$ as the digraph on n vertices obtained by including each ordered pair (u, v) as an arc independently and with probability p. Note that $D_{n,p}$ may have directed 2-cycles, i.e. it will contain both arcs (u, v) and (v, u) with probability p^2.

Our goal for this subsection is to use coupling arguments to show that $D_{n,p}$ and $G_{n,p}$ exhibit similar "behaviors", which will allow one to lift results from one setting to the other. Our approach can be generalized to a very large class of properties of graphs/digraphs, but for ease of presentation, we focus only on the property of Hamiltonicity. To this end, we say a digraph D is *Hamiltonian* if one can order the vertices as $v_1, \ldots, v_n$ such that $v_i v_{i+1} \in E(D)$ for all $1 \le i \le n$ (with indices being written cyclicly), and we will call such an ordering of its vertices a *Hamiltonian cycle*.

As a warmup, we prove the following basic coupling result.

Proposition 6.1. *For all $p \in [0, 1]$, we have*

$$\Pr[D_{n,p} \textit{ is Hamiltonian}] \le \Pr[G_{n,q} \textit{ is Hamiltonian}],$$

where $q := 2p - p^2$.

Proof. Our main goal is to construct (correlated) random variables D, G such that (1) D and G have the same distributions as $D_{n,p}$ and $G_{n,q}$, respectively, and such that (2) G is Hamiltonian whenever D is. From this the result will quickly follow. And in this case the path forward is relatively easy: start with a random digraph $D \sim D_{n,p}$, then take G to be the graph obtained from D by "forgetting" the orientations of each arc. One can easily check that conditions (1) and (2) are satisfied here. For completeness, we consider a more formal argument below.

Let $\{X_{u,v} : u, v \in [n]\}$ be a collection of iid Bernoulli random variables with success probability p. Define the random digraph D on $[n]$ by including the arc (u, v) iff $X_{u,v} = 1$ and define the random graph G on $[n]$ by including the edge uv if and only if $\max\{X_{u,v}, X_{v,u}\} = 1$. It is not difficult to see that D being Hamiltonian implies that G is Hamiltonian (since in particular, $v_1, \ldots, v_n$ being a Hamiltonian cycle in D implies it is also a Hamiltonian cycle in G), so in particular we find

$$\Pr[D \text{ is Hamiltonian}] \le \Pr[G \text{ is Hamiltonian}].$$

The result follows since (as is easy to check), D, G are random variables distributed according to $D_{n,p}, G_{n,q}$ respectively. $\square$

We now wish to reverse Proposition 6.1 by lowering bounding the probability that $D_{n,p}$ is Hamiltonian in terms of that for $G_{n,p}$ by using a very nice coupling by McDiarmid [116].

Theorem 6.2 ([116]). *For all $p \in [0,1]$, we have*

$$\Pr[G_{n,p} \textit{ is Hamiltonian}] \le \Pr[D_{n,p} \textit{ is Hamiltonian}].$$

Proof. Naively, one might try to perform the same approach as in Proposition 6.1 by starting with $G \sim G_{n,p}$ and then replacing each edge of G with some number of arcs. However, one can quickly work out that such a scheme will not recover a digraph distributed according to $D_{n,p}$, since in particular one expects the number of (unordered) pairs of vertices in $D_{n,p}$ connected by some arc to be about twice as many as the number of edges in $G_{n,p}$.

To remedy the situation, we use a clever idea of McDiarmid. Roughly speaking, instead of considering just the two models $G_{n,p}$ and $D_{n,p}$, we will instead consider a sequence of models $D_0, D_1, \ldots, D_N$ which "interpolate" between $G_{n,p}$ and $D_{n,p}$ in the sense that $D_0 = G_{n,p}$ and $D_N = D_{n,p}$ and are such that D_{i-1} and D_i differ only on how they include edges/arcs on a single pair of vertices e_i. Because there is such a small difference between the models D_{i-1} and D_i, we will easily be able to compare the probabilities of each being Hamiltonian, and by iterating this we will obtain our desired comparison for $D_0 = G_{n,p}$ and $D_N = D_{n,p}$.

As a small technical aside, it will be slightly more convenient in this argument to work exclusively with digraphs. To this end, we define a new random digraph model $D^*_{n,p}$ which independently for each unordered pair $\{u,v\}$ includes both arcs $(u,v),(v,u)$ with probability p and excludes both arcs with probability $1-p$. That is, $D^*_{n,p}$ is equivalent to the digraph obtained from $G_{n,p}$ by replacing each edge uv with the 2-cycle $(u,v),(v,u)$, and in particular the probability that $G_{n,p}$ is Hamiltonian is equal to the probability that $D^*_{n,p}$ is Hamiltonian. As such, it suffices to work with this latter model instead.

We now turn to the formal details. Let $e_1, \ldots, e_{\binom{n}{2}}$ be an arbitrary ordering of the unordered pairs of $[n]$, say with $e_j = \{u_j, v_j\}$. For all $0 \le i \le \binom{n}{2}$, define the random digraph D_i on $[n]$ as follows: for all $j > i$, independently and with probability p include both of the arcs (u_j, v_j) and (v_j, u_j) in D_i; and for all $j \le i$, independently and with probability p include the arc (u_j, v_j), and independently and with probability p include the arc (v_j, u_j). Observe that $D_0 \sim D^*_{n,p}$ and that $D_{\binom{n}{2}} \sim D_{n,p}$ (with more generally D_i having exactly i pairs behaving like those in $D_{n,p}$ and the rest like $D^*_{n,p}$), so we will be done if we can show for all $i \ge 1$ that

$$\Pr[D_{i-1} \text{ is Hamiltonian}] \le \Pr[D_i \text{ is Hamiltonian}].$$

It is not difficult to couple D_{i-1} and D_i in such a way that they agree outside of the pair e_i, and we let D denote this common digraph. Let A denote the event that D is not Hamiltonian but adding at least one arc of e_i makes D Hamiltonian. Observe that

$$\Pr[D_{i-1} \text{ is Hamiltonian} | A^c] = \Pr[D_i \text{ is Hamiltonian} | A^c],$$

since either D is Hamiltonian (in which case D_{i-1}, D_i will both be Hamiltonian with probability 1), or D plus both arcs of e_i is not Hamlitonian (in which case D_{i-1}, D_i will both be Hamiltonian with probability 0). Conditional on A, we have that $\Pr[D_{i-1} \text{ is Hamiltonian} | A^c] = p$, as we will be Hamiltonian iff we include both arcs on e_i. Crucially, we also have $\Pr[D_i \text{ is Hamiltonian} | A^c] \ge$

p, as p is the probability of including some given arc on e_i which will make D_i Hamiltonian (with it being possible for two such arcs to exist, in which case the probability becomes higher). Thus regardless of the event A we see that D_i is at least as likely to be Hamiltonian as D_{i-1}, proving the claim and hence the proof. $\square$

This coupling approach is very flexible and has found applications in many other settings

Discuss variants of this method, e.g. to rainbow setting as well as perturbed random setting. See e.g. "Rainbow subgraphs of uniformly coloured randomly perturbed graphs", "Rainbow Hamilton cycles in random graphs and hypergraphs", and " Spanning cycles in random directed graphs,"

6.2 Graphs and Hypergraphs

Here we discuss a beautiful coupling due to Riordan and Heckel allowing us to translate facts about random hypergraphs to random graphs. As a partial warmup to the argument, we consider a coupling result in a completely unrelated setting (namely that of random walks), whose proof is spiritually similar to theirs while having the advantage of being substantially simpler.

Let S^n denote a simple random walk of length n, i.e. S^n is a random vector $(S_0^n, S_1^n, \dots, S_n^n)$ where $S_0^n = 0$, and $\Pr[S_i^n = S_{i-1}^n + 1] = \Pr[S_i^n = S_{i-1}^n - 1] = \frac{1}{2}$. For n even, let T^n denote a random walk after conditioning on having $T_n^n = 0$ (i.e. we uniformly at random pick a walk which returns to to 0 at the end of the walk). It is easy to show via Chernoff boudns that S_t^n is likely to be within roughly $\sqrt{t}$ of 0 for any given value t. While Chernoff bounds don't apply to the random variables T_t^n, intuitively the same conclusion should also hold for T_t^n, since the condition of $T_n^n = 0$ should force T_t^n to be closer to the origin than S_t^n in general. It is possible to make this intuition rigorous, allowing one to bootstrap bounds of S_t^n to T_t^n.

Proposition 6.3. *For n even and all s, t, we have* $\Pr[|T_t^n| \geq s] \leq \Pr[|S_t^n| \geq s]$.

Proof. Our goal is to define a new random vector R^n such that (1) R^n has the same distribution as T^n, and (2) $|R_t^n| \leq |S_t^n|$ for all t. From this the result will quickly follow. Intuitively, we will define R_t^n in rounds by flipping biased coins. If the tth coin lands heads, then R_{t+1}^n moves towards 0, and if it lands tails, it moves towards/away from 0 if and only if S_{t+1}^n moves towards/away from 0. Such a process will always satisfy (2), and it will satisfy (1) by choosing the probability of our biased coins appropriately.

To this end, set $R_0^n = 0$. Given R_t^n, we define a random variable Y_t (which will be our biased coin flips) that equals 1 with probability $\frac{|R_t^n|}{n-t}$ and is 0 otherwise. If $Y_t = 1$, we set $R_{t+1}^n = R_t^n \pm 1$ such that $|R_t^n| > |R_{t+1}^n|$ (i.e. such that R^n moves towards 0; note that this is well defined since $Y_t = 1$ implies $R_t^n \neq 0$). If $Y_t = 0$ and $R_t^n \neq 0$, then we set $R_{t+1}^n = R_t^n \pm 1$ such that $|R_t^n| > |R_{t+1}^n|$ if and only if $|S_t^n| > |S_{t+1}^n|$ (i.e. R^n move away/towards 0 if S^n moves away/towards 0). If $R_t^n = 0$ then we set $R_t^n = \pm 1$ with equal probability.

It is straightforward to see that (2) is achieved from this process[1]. It is not difficult to prove that

[1]Any time S^n moves towards 0, R^n does as well, except when $R_t^n = 0$. In this case S_t^n must be an even distance away from 0, so after one step R_t^n is still at least as close to 0.

for T^n, we have that $|T^n_t| > |T^n_{t+1}|$ happens with probability[1] $\frac{\frac{1}{2}(n-t+|T^n_t|)}{n-t}$. One can check that R^n has $|R^n_t| > |R^n_{t+1}|$ with probability $\frac{\frac{1}{2}(n-t+|R^n_t|)}{n-t}$, so we conclude (1) and hence the result. $\square$

Recall that an F-*factor* in a graph G is a collection of vertex disjoint copies of F such that every vertex is in one of these copies of F. By a similar argument as in Theorem 8.2, one can show that $G_{n,p}$ contains a K_r-factor provided $r|n$ and $p \gg n^{-1/\binom{r}{2}} \log n$. Intuitively, it seems reasonable that the set of K_r's in $G_{n,p}$ should be distributed like the hyperedges of $G^r_{n,\pi}$ where $\pi = p^{\binom{r}{2}}$ (at the very least, the expected number of K_r's in $G_{n,p}$ is equal to the expected number of hyperedges in $G^r_{n,\pi}$). If this were true, then $G_{n,p}$ would contain a K_r-factor when $G^r_{n,\pi}$ contains a perfect matching, and Theorem 8.2 says this should happen when $p^{\binom{r}{2}} \approx \pi \gg n^{-1} \log n$, which implies that taking $p \gg n^{-1/\binom{r}{2}} \log^{1/\binom{r}{2}} n$ should suffice. And indeed, Johansson, Kahn, and Vu [95] proved that this is the threshold for K_r-factors in $G_{n,p}$ using a somewhat involved argument. A nice coupling result of Riordan [129] will allow us to conclude the result in a much easier way.

Let $(V_1, E_1), \ldots, (V_{\binom{n}{r}}, E_{\binom{n}{r}})$ be an arbitrary ordering of all the K_r's in K_n. To prove our desired coupling, we would like to construct a pair of random variables (G, H) such that (1) $G \sim G_{n,p}$ and $H \sim G^r_{n,\pi}$ with $\pi \approx p^{\binom{r}{2}}$, and such that (2) every hyperedge in H is a K_r in G. Note that (2) means that H containing a perfect matching implies that G has a K_r-factor. Let us first consider the following (very, very) naive attempt at this coupling.

Algorithm 1. Generate a random graph $G \sim G_{n,p}$. Let H be an initially empty r-graph on $[n]$. For each i with $E_i \subseteq G$, add V_i as a hyperedge to H. Output (G, H).

This algorithm definitely satisfies (2), but it completely fails at (1). Indeed, let A_i denote the event that $E_i \subseteq G$ (i.e. the event that V_i is a hyperedge in H), and assume V_1, V_2 have at least two vertices in common. Then $\Pr[A_2|A_1] \geq p^{\binom{r}{2}-1}$ and $\Pr[A_2|\overline{A_1}] < p^{\binom{r}{2}}$. But to have $H \sim G^r_{n,\pi}$ we would, in particular, need these two probabilities to equal each other. Thus we'll need to consider a somewhat more complicated algorithm. As before, let (V_i, E_i) be the K_r's in K_n, and let π be a parameter which will be approximately $p^{\binom{r}{2}}$.

Algorithm 2. Generate a random graph $G \sim G_{n,p}$ and an initially empty hypergraph H on $[n]$. We proceed in $\binom{n}{r}$ rounds as follows. For the ith round, let π_i be the conditional probability of having $E_i \subseteq G$ given all the information from the previous rounds.

- If $\pi_i < \pi$, then with probability π we add V_i to H.
- If $\pi_i \geq \pi$, then with probability $\frac{\pi}{\pi_i}$ we test whether $E_i \subseteq G$, and if so, we add V_i to H. Otherwise[2] we declare this hyperedge to be absent in H.

[1] Given T^n_t, we still need to make $\frac{1}{2}(n-t+|T^n_t|)$ steps in the direction of 0 and $\frac{1}{2}(n-t-|T^n_t|)$ in the direction away from 0.

[2] Note that with probability $1 - \frac{\pi}{\pi_i}$ we do not reveal any additional information about E_i. When working with random objects, it is usually best to reveal as little information as possible in order to "preserve" the randomness of your object.

We note that the $\pi_i \geq \pi$ case of Algorithm 2 is similar in spirit to the proof of Proposition 6.3: each round we flip a coin which is biased based off of the current information we have. If the coin lands heads we do something to H independent of G, and otherwise we have H behave "in the same way" as G.

For Algorithm 2, it is not difficult to see that $H \sim G^r_{n,\pi}$. Unfortunately, if $\pi_i < \pi$, then it is possible that H contains edges which are not K_r's in G, i.e. the coupling could fail to satisfy (2). The key insight is that for applications, it suffices to have (2) be satisfied with high probability, which will turn out to be the case.

To try and convince ourselves that this algorithm has a chance of winning even when $\pi_i < \pi$, let's consider the most dangerous situation, namely that $\pi_i = 0$. It is not too hard to see that $\pi_i = 0$ if and only if there exists some $j < i$ such that (a) we revealed that $E_j \not\subseteq E(G)$ and (b) every edge of $E_j \setminus E_i$ has been revealed to be in G. If this situation happens and if the algorithm adds V_i to H, then the coupling fails to satisfy (2). However, when this happens, every graph edge of E_j is contained in a hyperedge of H (by (b) and $V_i \in E(H)$) and V_j is not a hyperedge of H (by (a)). Thus the probability of this situation happening is at most the probability of $H \sim G^r_{n,\pi}$ containing such a configuration. These configurations can essentially be described as follows.

Lemma 6.4. *If H is an r-graph with $r \geq 4$ which contains a set of r vertices $V \notin E(H)$ such that every pair of V is contained in a hyperedge of H, then H contains a subgraph F which has $e(F) \leq \binom{r}{2}$ and $|V(F)| \leq (r-1)e(F) - 1$.*

This statement is false for $r = 3$. Indeed, one could take H to be the loose triangle with edges $\{1, 2, 4\}, \{2, 3, 5\}, \{1, 3, 6\}$ which satisfies the hypothesis of Lemma 6.4 with $V = \{1, 2, 3\}$ but which fails to satisfy the conclusion.

Proof. Let $V_1, \ldots, V_t$ be hyperedges such that every pair of V is contained in some V_i. By throwing away redundant hyperedges, we can assume that $|V_i \cap V| \geq 2$ for all i and that $t \leq \binom{r}{2}$. Let $F \subseteq H$ be the hypergraph with hyperedges $V_1, \ldots, V_t$

First assume $|V_1 \cap V_2| \geq 2$. Then $V_1 \cup V_2$ consists of at most $2r - 2$ vertices, and it is not difficult to see that it is possible to order the remaining sets so that $|V_i \setminus \bigcup_{j<i} V_j| \leq r - 1$ and that $|V_t \setminus \bigcup_{j<t} V_j| \leq r - 2$. In total this implies that F has the desired properties.

Thus we can assume that $|V_i \cap V_j| \leq 1$ for all i, j. This means every pair of V is covered by some unique V_i, so $t = e(F) = \binom{r}{2}$ and the number of vertices of F is at most $r + (r-2)e(F) = (r-1)e(F) - (e(F) - r) \leq (r-1)e(F) - 1$ since $\binom{r}{2} - r \geq 1$ for $r \geq 4$. $\square$

Lemma 6.5. *For $r \geq 4$, if $H \sim G^r_{n,\pi}$ and $\pi \leq n^{-(r-1)+o(1)}$, then a.a.s. H does not contain a set V as in Lemma 6.4.*

Proof. If H did contain such a set V, then it must contain a subgraph F as in Lemma 6.4. Up to isomorphism, there are only finitely many subgraphs that F could be, and for each of these the expected number of copies of F in H is at most

$$O(\pi^{e(F)} n^{|V(F)|}) = O(\pi^{e(F)} n^{(r-1)e(F)-1}) = o(1).$$

We conclude the result by Markov's inequality. $\square$

We note that for $\pi \approx p^{\binom{r}{2}}$ this lemma applies when $p \approx n^{-2/r}$. Thus when p is about this value, none of the "bad" configurations of Lemma 6.4 are likely to appear, and in this regime we have the following.

Theorem 6.6 ([129])**.** *For $r \geq 4$ and $p \leq n^{-2/r+o(1)}$, there exists some $\pi \sim p^{\binom{r}{2}}$ such that Algorithm 2 produces a pair (G, H) with $G \sim G_{n,p}$, $H \sim G^r_{n,\pi}$, and such that a.a.s. every hyperedge of H is the vertex set of a K_r in G.*

We emphasize that the theorem as stated does not cover the case $r = 3$. However, Heckel [87] showed that the same conclusion does hold for $r = 3$ by using a slightly different coupling.

For the proof of Theorem 6.6, we will need a standard result known as Harris' inequality (also referred to as Kleitman's inequality).

Lemma 6.7 (Harris' Inequality)**.** *Let $f, g, h : \mathbb{R}^n \to \mathbb{R}$ be functions such that f, g are non-decreasing and h is non-increasing. Let $X = (X_1, \ldots, X_n)$ be a random vector such that the X_i's are mutually independent. Then*

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)]\mathbb{E}[g(X)],$$

$$\mathbb{E}[f(X)h(X)] \leq \mathbb{E}[f(X)]\mathbb{E}[h(X)].$$

Proof. For $n = 1$, we deterministically have

$$(f(y) - f(z))(g(y) - g(z)) \geq 0 \geq (f(y) - f(z))(h(y) - h(z)).$$

Thus if Y, Z are independent random variables with the same distribution as $X = X_1$, the first inequality implies

$$0 \leq \mathbb{E}[f(Y)g(Y) + f(Z)g(Z) - f(Y)g(Z) - f(Z)g(Y)] = 2\mathbb{E}[f(X)g(X)] - 2\mathbb{E}[f(X)]\mathbb{E}[g(X)].$$

This gives the first bound, and the second bound follows from an identical argument.

Assume the result has been proven up to some $n > 1$. By the inductive hypothesis and the $n = 1$ case applied to $f'(X_1) := \mathbb{E}[f(X)|X_1]$ and $g'(X_1) = \mathbb{E}[g(X)|X_1]$, we find

$$\begin{aligned}\mathbb{E}[f(X)g(X)] &= \mathbb{E}[\mathbb{E}[f(X)g(X)|X_1]] \geq \mathbb{E}[\mathbb{E}[f(X)|X_1] \cdot \mathbb{E}[g(X)|X_1]] \\ &= \mathbb{E}[f'(X_1)g'(X_1)] \geq \mathbb{E}[f'(X_1)]\mathbb{E}[g'(X_1)] = \mathbb{E}[f(X)]\mathbb{E}[g(X)].\end{aligned}$$

This proves the first inequality, and the second follows from an identical argument. ☐

The main application of Harris' inequality is when f is an indicator function. More precisely, we say that a set system $\mathcal{A} \subseteq 2^{[n]}$ is an *upset* if $A \in \mathcal{A}$ implies $B \in \mathcal{A}$ for all $B \supseteq A$, and we similarly define what it means for $\mathcal{A}$ to be a *downset*.

Corollary 6.8. *Let $\mathcal{A}, \mathcal{B}$ be upsets and $\mathcal{C}$ a downset of $[n]$, and let $S \subseteq [n]$ be obtained by including each element i independently and with probability p_i. Then*

$$\Pr[S \in \mathcal{A} \cap \mathcal{B}] \geq \Pr[S \in \mathcal{A}] \Pr[S \in \mathcal{B}],$$

$$\Pr[S \in \mathcal{A} \cap \mathcal{C}] \leq \Pr[S \in \mathcal{A}] \Pr[S \in \mathcal{C}].$$

Proof. Define $f : \mathbb{R}^n \to \mathbb{R}$ by having $f(x) = 1$ if $\{i : x_i > 0\} \in \mathcal{A}$ and $f(x) = 0$ otherwise. Similarly define g, h with respect to $\mathcal{B}, \mathcal{C}$. The result follows from Harris' inequality by letting $X = (X_1, \ldots, X_n)$ with the X_i being independent Bernoulli random with probability p_i. $\square$

Proof of Theorem 6.6. Again, the only cases where the algorithm can fail is when π_i is small, so let us try and lower bound this quantity in terms of the (random) information we have at the ith step. Let Y be the set of "yes" indices j such that we have revealed that $E_j \subseteq E(G)$ and let N be the set of "no" indices such that we have revealed $E_j \not\subseteq E(G)$. Let $R = \bigcup_{j \in Y} E_j$ be the set of revealed edges, and let G' be the random graph which contains all of the edges of R, and which contains any $e \notin R$ independently and with probability p. Let $E'_j = E_j \setminus R$, and let A'_j be the event that $E'_j \subseteq G'$. It is not too hard to see that in total we have

$$\pi_i = \Pr[A'_i | \bigcap_{j \in N} \overline{A'_j}].$$

Define

$$D_0 = \bigcap_{j \in N,\ E_j \cap E_i = \emptyset} \overline{A'_j}, \qquad D_1 = \bigcap_{j \in N,\ E_j \cap E_i \neq \emptyset} \overline{A'_j}.$$

Intuitively D_0 shouldn't really influence π_i, and we can prove this using Harris' inequality. First note that

$$\pi_i = \Pr[A'_i | D_0 \cap D_1] \geq \Pr[A'_i \cap D_1 | D_0] = \Pr[A'_i | D_0] - \Pr[A'_i \cap \overline{D_1} | D_0] = \Pr[A'_i] - \Pr[A'_i \cap \overline{D_1} | D_0],$$

where this last step used that A'_i and D_0 are independent. Observe that A'_j is an upset for all j (i.e. A'_j is achieved precisely when the random set $E(G')$ is an element of an appropriately defined upset), D_0 is a downset (since complements of upsets are downsets, and downsets/upsets are preserved under intersection), and $A'_i \cap \overline{D_1}$ is an upset. Thus by Harris' inequality, we have[1]

$$\Pr[A'_i] - \Pr[A'_i \cap \overline{D_1} | D_0] = \Pr[A'_i] - \frac{\Pr[A'_i \cap \overline{D_1} \cap D_0]}{\Pr[D_0]} \geq \Pr[A'_i] - \Pr[A'_i \cap \overline{D_1}].$$

Now let $N_1 = \{j \in N : E_j \cap E_i \neq \emptyset\}$. Note that $\overline{D_1} = \bigcup_{j \in N_1} A'_j$, so by a union bound we have

$$\pi_i \geq \Pr[A'_i] - \sum_{j \in N_1} \Pr[A'_j \cap A'_i] = p^{|E'_i|} - \sum_{j \in N_1} p^{|E'_i \cup E'_j|} = p^{|E'_i|}(1 - Q_i) \geq p^{\binom{r}{2}}(1 - Q_i),$$

where

$$Q_i := \sum_{j \in N_1} p^{|E_j \setminus (E_i \cup R)|}.$$

Let Δ denote the maximum degree of G. We next prove a (somewhat imprecise) claim.

Claim 6.9. *Either $\Delta > n^{o(1)}$, or for all i, either $Q_i = o(1)$ or $V_i \in E(H)$ implies H contains a configuration as in Lemma 6.4.*

[1] This same sort of argument is essentially what you need to do prove Janson's inequality Theorem 3.7.

Proof. Assume $\Delta \le n^{o(1)}$ and consider some index i. Given j, let K_j denote the graph on V_j with edge set $E_i \cup R$, and let $C_1, \ldots, C_{k+1}$ with $k \ge 0$ denote the connected components of K_j, say with $|V(C_\ell)| = r_\ell$ for all ℓ. Observe that $|E_j \setminus (E_i \cup R)|$ is at least the number of edges which aren't contained in any K_j component, i.e.

$$|E_j \setminus (E_i \cup R)| \ge \binom{r}{2} - \sum \binom{r_\ell}{2} \ge \binom{r}{2} - \binom{r-k}{2},$$

where this last inequality holds since if there are two terms with $r_\ell \ge 2$, then one can adjust these two terms to get a stronger bound. Because K_j is a graph using edges of $R \cup E_i \subseteq G$, we have that the number of $j \in N_1$ such that K_j has $k+1$ components is at most $rn^k \Delta^{r-k-1} = n^{k+o(1)}$, where the factor of r comes from the fact that $j \in N_1$ implies that K_j contains at least one vertex of V_i since E_i, E_j intersect in at least one edge.

In total then, the contribution to Q_i coming from j such that K_j has $k+1 \ge 2$ components is at most

$$\sum_{k=1}^{r-2} n^{k+o(1)} p^{\binom{r}{2} - \binom{r-k}{2}} = o(1),$$

where the equality follows from a simple calculation. Thus it remains to show that the contribution from terms with K_j connected is small. Because there are only $r\Delta^{r-1} = n^{o(1)}$ such terms, a similar argument shows that the contribution is negligible for terms with $e(K_j) < \binom{r}{2}$. The only non-trivial case then is when $e(K_j) = \binom{r}{2}$, i.e. when every edge of V_j is contained in $R \cup E_i$. In this case, $V_i \in E(H)$ implies that H contains a configuration as in Lemma 6.4. Thus for all i, either this happens or $Q_i = o(1)$, proving the result. □

Since $H \sim G^r_{n,\pi}$, we have that the expected degree of every vertex is roughly $\pi n^{r-1} = n^{o(1)}$. Thus if $\mathcal{B}_1$ is the "bad" event that $\Delta > n^{o(1)}$, then by the Chernoff bound we have $\Pr[\mathcal{B}_1] = o(1)$. Similarly if $\mathcal{B}_2$ is the event that H contains one of the configurations as in Lemma 6.4, then $\Pr[\mathcal{B}_2] = o(1)$ by Lemma 6.5.

We are now ready to complete the proof. Recall that the theorem claims the result holds for *some* $\pi \sim p^{\binom{r}{2}}$, so it suffices to prove it for $\pi = p^{\binom{r}{2}}(1 - o(1))$ where the $o(1)$ term is the upper bound for Q_i from the claim. Now all we have to do is verify that with this choice, a.a.s. every hyperedge of H is a K_r in G. The only way this can fail is if there exists an i with $\pi_i < \pi$ such that V_i is added as a hyperedge to H. By the previous claim and our choice of π, this is only possible if $\mathcal{B}_1 \cup \mathcal{B}_2$ occurs. As these occur with probability $o(1)$, we conclude the result. □

As noted previously, the proof of Theorem 6.6 does not go through for $r = 3$ due to the existence of loose triangles, but Heckel [87] managed to get around this issue. Essentially the idea of her proof is to first do a coupling on edges of H and G which are in loose triangles and then to run Riordan's argument.

It is also proven in [129] that one can to some extent generalize this approach to finding F-factors for sufficiently nice F. In this setting, H is not exactly a uniform hypergraph, but instead a collection of copies of F in K_n chosen with some probability π.

7 Random Algebraic Constructions

One can easily extend our general lower bound for graph Turán numbers Theorem 1.4 to the setting of hypergraphs as follows.

Theorem 7.1. *Let F be an r-graph with v vertices and $e \geq r$ edges. If $e \geq v$, then*

$$\mathrm{ex}(n, F) = \Omega_v(n^{r - \frac{v-r}{e-1}}).$$

Sketch of Proof. Consider $G^r_{n,p}$, which in expectation has about pn^r edges and $p^{e(F)}n^{|V(F)|}$ copies of F. At $p = Cn^{-\frac{v-r}{e-1}}$ for some large constant C this first quantity is much larger than the second, so we can delete an edge from each copy of F to give the result. $\square$

One way you could try and improve upon this argument is to delete edges which are in many copies of F. In $G^r_{n,p}$ this is too much to ask for, but it is possible to do this in other random hypergraph models. In particular, if our random model contains some algebraic structure, then it is often the case that edges will either be in many copies of F or almost none. We look at a few examples of this phenomenon.

7.1 Random Multilinear Maps

The problem of determining the Turán number of $K^r_{2,\ldots,2}$, the complete r-partite r-graph with each part having size 2, is called the Erdős box problem. Theorem 7.1 gives a lower bound of $n^{r-\frac{r}{2^r-1}}$, and for certain values of r this lower bound was improved by Gunderson, Rödl, and Sidorenko [85]. This result was significantly improved by Conlon, Pohoata, and Zakharov[42] who gave a polynomial improvement to the bound of Theorem 7.1 for all values of r.

Theorem 7.2 ([42])**.** *For all $r \geq 2$, we have*

$$\mathrm{ex}(n, K^r_{2,\ldots,2}) = \Omega(n^{r - \lceil \frac{2^r-1}{r} \rceil^{-1}}).$$

Note that r never[1] divides $2^r - 1$, so this does always give a polynomial improvemnt to Theorem 1.4.

We prove this result by considering a random hypergraph based off of multilinear maps. Recall that if $V_1, \ldots, V_r$ are vector spaces over $\mathbb{F}_q$, then a map $T : V_1 \times \cdots \times V_r \to \mathbb{F}_q$ is said to be *multilinear* if the one dimensional function $f(x) = T(v_1, \cdots, v_{i-1}, x, v_{i+1}, \cdots, v_r)$ is linear for all i and any choice of v_j. Note that there are only finitely many such maps over $\mathbb{F}_q$ if $V_1, \ldots, V_r$ are finite dimensional, so in this setting we can talk about choosing such a T uniformly at random.

Let $s = \lceil \frac{2^r-1}{r} \rceil$, and let $V_1, \ldots, V_r$ be copies of $\mathbb{F}^s_q$ with q a large prime power. Given a multilinear map T, let H_T denote the r-partite r-graph on $V_1 \cup \cdots \cup V_r$ with $\{v_1, \ldots, v_r\} \in E(H_T)$ if and only if $T(v_1, \ldots, v_r) = 1$, where here and throughout we assume $v_i \in V_i$ for all i. The proof relies on the following three results.

[1] If r is prime then $2^r - 1 \equiv 2 - 1 \mod r$ by Fermat's little theorem **though I don't see why this holds otherwise**

Lemma 7.3. *Let T be a uniformly random multilinear map and assume q is sufficiently large in terms of r. Then the following hold:*

(a) We have $\mathbb{E}[e(H_T)] = (q^s - 1)^r q^{-1} \approx q^{rs-1}$.

(b) Let $\mathcal{F}$ denote the set of tuples $(v_1^0, v_1^1, \dots, v_r^0, v_r^1)$ with $v_i^j \in V_i$ and $v_i^0 \neq v_i^1$ such that $T(v_1^{j_1}, \dots, v_r^{j_r}) = 1$ (i.e. such that this forms a $K^r_{2,\dots,2}$ in H_T). Then $\mathbb{E}[|\mathcal{F}|] \sim q^{2rs-2^r}$.

(c) Let $\mathcal{B}$ denote the set of edges $\{v_1, \dots, v_r\}$ such that $(v_1, v_1', \dots, v_r, v_r') \in \mathcal{F}$ for some $\{v_1', \dots, v_r'\}$. Then $\mathbb{E}[|\mathcal{B}|] \leq (1 + o(1))q^{-r}\mathbb{E}[|\mathcal{F}|]$.

We note that $G^r_{n,p}$ with p, n chosen appropriately already roughly satisfy (a) and (b), so the crucial thing we gain here is (c), which says that there are not many edges that are contained in some $K^r_{2,\dots,2}$, i.e. the copies of $K^r_{2,\dots,2}$ are all clumped together. This is the key fact that we acquire from using a random algebraic construction.

Let us briefly observe that this lemma gives the result. Indeed, we can form a $K^r_{2,\dots,2}$-free hypergraph H_T' by deleting every edge of $\mathcal{B}$. The expected number of edges for this will be asymptotically at least $q^{rs-1} - q^{2rs-2^r-r}$. Because $s = \lceil \frac{2^r-1}{r} \rceil$, we have $s < \frac{2^r-1}{r} + 1$, which is equivalent to saying $rs - 1 > 2rs - 2^r - r$, and hence the number of edges is roughly q^{rs-1}. Since H_T' has $rq^s := n$ vertices, this gives $\mathrm{ex}(n, K^r_{2,\dots,2}) = \Omega(n^{r-1/s})$ as desired. Thus it remains to prove the lemma.

Proof of Lemma 7.3. For (a), note that $T(v_1, \dots, v_r) = 0$ if $v_i = 0$ for some i. For any other tuple, let $U_i \subseteq V_i$ be the one-dimensional subspace containing v_i and 0. It is not too hard to argue that T restricted to $U_1 \times \cdots \times U_r$ is still a uniform multilinear map. Further, every multilinear map on this space is uniquely determined by the value of $T(1, \dots, 1)$, and it is not hard to see that exactly one of these q maps has $T(v_1, \dots, v_r) = 1$. Thus such a tuple is an edge with probability q^{-1} and the result follows from linearity of expectation.

For (b), observe that the only tuples that can be in $\mathcal{F}$ are those such that $v_i^0 \neq \lambda v_i^1$ for any i, as otherwise

$$\lambda = \lambda T(v_1^0, \dots, v_i^0, \dots, v_r^0) = T(v_1^0, \dots, v_i^1, \dots, v_r^0) = 1,$$

which means $\lambda = 1$, contradicting $v_i^0 \neq v_i^1$. The number of such tuples with this property is asymptotic to q^{2rs}. For such a tuple, let U_i be the span of v_i^0, v_i^1, which is a 2-dimensional subspace. Again T restricted to $U_1 \times \cdots \times U_r$ is uniform, and it is not too hard to see that there are q^{2^r} choices for T with exactly one of these placing the tuple in $\mathcal{F}$. The result follows from linearity of expectation.

It remains to deal with (c). Given affine lines $\ell_1, \dots, \ell_r$ in $V_1, \dots, V_r$, let $P(\ell_1, \dots, \ell_r)$ denote the set of tuples $(v_1, v_1', \dots, v_r, v_r')$ such that $v_i, v_i' \in \ell_i$. It is not difficult to show the following:

- The sets $P(\ell_1, \dots, \ell_r)$ are disjoint for distinct choices of lines.
- We have $|P(\ell_1, \dots, \ell_r)| = q^r(q-1)^r$.
- Every element of $\mathcal{F}$ is contained in some $P(\ell_1, \dots, \ell_r)$.

- If $P(\ell_1, \ldots, \ell_r) \cap \mathcal{F} \neq \emptyset$ then $P(\ell_1, \ldots, \ell_r) \subseteq \mathcal{F}$, i.e. $T(u_1, \ldots, u_r) = 1$ for any $u_i \in \ell_i$.

If $\mathcal{L}$ denotes the set of tuples $(\ell_1, \ldots, \ell_r)$ with $P(\ell_1, \ldots, \ell_r) \cap \mathcal{F} \neq \emptyset$, then the above implies that

$$|\mathcal{L}|q^r(q-1)^r = |\mathcal{F}|.$$

Further, we have

$$|\mathcal{B}| = \left| \bigcup_{(\ell_1,\ldots,\ell_r)\in\mathcal{L}} \ell_1 \times \cdots \times \ell_r \right| \le q^r|\mathcal{L}| = (q-1)^{-r}|\mathcal{F}|,$$

so taking expectations gives the result. □

We note that one can get a slightly stronger result by not just considering one multilinear map T, but a family of (random) multilinear maps $T_1, \ldots, T_\ell$ and then defining $H_{T_1,\ldots,T_\ell}$ by having a hyperedge if and only if $T_i(v_1, \ldots, v_r) = 1$ for all i. The analysis here is mostly the same, but for ease of presentation we only considered a single map.

7.2 Random Polynomial Graphs

Somewhat more complicated constructions can be made by utilizing random polynomials as opposed to random multilinear maps. This approach was first popularized by Bukh [30], and since then Bukh and Conlon have developed a lot of theory surrounding it.

To set things up, given a field $\mathbb{F}_q$, we define $\mathcal{P}_{d,b}$ to be the set of polynomials over $\mathbb{F}_q$ in t variables with degree at most d. We will say that f is a random polynomial from $\mathcal{P}_{d,b}$ if it is chosen uniformly at random from $\mathcal{P}_{d,b}$, which can be done, for example, by uniformly at random choosing the coefficient of each possible monomial. With a little bit of linear algebra one can show the following, which says that a random polynomial has the same distribution as a random function when evaluated on a few number of points.

Lemma 7.4 ([32] Lemma 2.3). *If $q > \binom{m}{2}$ and $d \geq m-1$, then if $f \in \mathcal{P}_{d,b}$ is uniformly random and $x_1, \ldots, x_m$ are m distinct points of $\mathbb{F}_q^b$, then*

$$\Pr[f(x_i) = 0 \ \ \forall i] = q^{-m}.$$

Maybe include proof.

The next lemma requires just a smidge of terminology from algebraic geometry. A *variety* is any set of the form $X = \{x \in \mathbb{F}_q^b : f_1(x) = \cdots = f_a(x) = 0\}$ where $f_1, \ldots, f_a : \mathbb{F}_q^t \to \mathbb{F}_q$ are polynomials. The variety X is said to have complexity at most M if a, b and the degrees of the f_i are bounded by M. One can prove the following using standard results from algebraic geometry.

Lemma 7.5 ([32] Lemma 2.7). *Let X, D be varieties over $\mathbb{F}_q$ of complexity at most M. If q is sufficiently large in terms of M, then either $|X \setminus D| \geq q/2$ or $|X \setminus D| \leq c$ for some c depending only on M.*

The actual lemma statement involves the algebraic closure, which I think is an artifact of the proof and isn't necessary in the statement. Please let me know if you think I'm wrong (or right) about this point.

One can think of this lemma as being analogous to the fact that if f is a degree d polynomial in one variable which is 0 on at least $d+1$ points, then it must in fact be 0 on an entire line. With these two results we can prove the following.

Theorem 7.6. *For all $s \geq 2$, there exists some $t_0 = t_0(s)$ such that for all $t \geq t_0$, we have*

$$\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s}).$$

Proof. The upper bound follows from the Kővári-Sós-Túran theorem. For the lower bound, let q be a sufficiently large prime power, and with some foresight we define

$$r = s^2 + 1, \quad d = rs + 1, \quad N = q^s.$$

Let $f \in \mathcal{P}_{d,2s}$ be a polynomial chosen uniformly at random. Let G be the (random) graph with vertex set $\mathbb{F}_q^s \times \mathbb{F}_q^s$ where vertices $x^1 \in \mathbb{F}_q^s, x^2 \in \mathbb{F}_q^s$ form an edge of G if and only if $f(x^1, x^2) = 0$.

Fix vertices $x^1, \ldots, x^s \in \mathbb{F}_q^s \cup \mathbb{F}_q^s$. Let C be the set of vertices y such that $x^i \sim y$ for all i (noting that $C = \emptyset$ if the x^i don't all belong to the same copy of $\mathbb{F}_q^s$, and otherwise this means e.g. $f(x^i, y) = 0$ for all i). Observe that the number of $K_{s,r}$'s of G which has the x^i as its set of size s is equal to $\binom{|C|}{r}$, and motivated by this we will attempt to bound the rth moment $\mathbb{E}[|C|^r] = \mathbb{E}[|C^r|]$ (which will be slightly easier to work with compared to the rth falling moment). To this end, we observe that if a given tuple $(y^1, \ldots, y^r)$ with k distinct elements lies in C^r, then the corresponding copy of $K_{s,k}$ lies in G. By Lemma 7.4, the probability that any given copy of $K_{s,k}$ appears in G is exactly q^{-sk} (provided q is sufficiently large in terms of s, r). Moreover, the number of tuples with k distinct elements is $O_r(N^k)$. In total we conclude that **The exposition here can probably be cleaned up**

$$\mathbb{E}[|C^r|] \leq \sum_{k=1}^{r} q^{-sk} \cdot O_r(N^k) = O_r(1).$$

Note that C is an algebraic variety by definition. By Lemma 7.5, there exists some constant c such that either $|C| \leq c$ or $|C| \geq q/2$. Thus

$$\Pr[|C| > c] = \Pr[|C| \geq q/2] = \Pr[|C|^r \geq (q/2)^r] \leq \frac{\mathbb{E}[|C|^r]}{(q/2)^r} = O_r(q^{-r}),$$

with the last step using the previous inequality.

Call a sequence $(x^1, \ldots, x^s)$ *bad* if there are more than c vertices y such that $x^i \sim y$ for all i, and let B_i denote the number of i-bad sequences. Our analysis above gives

$$\mathbb{E}[B_i] \leq 2N^s \cdot O_r(q^{-r}) = O_r(q^{s^2 - r}) = o_r(1). \tag{8}$$

Now let $G' \subseteq G$ be defined by deleting a vertex from each bad sequence. Because each vertex is in at most $N = q^s$ edges in G, by (**??**) and (8) we find

$$\mathbb{E}[e(G')] \geq \mathbb{E}[e(G)] - \mathbb{E}[B] \cdot q^s = \Omega(q^{2s-1}),$$

where this last step used the previous inequality and Lemma 7.4 to deduce $\mathbb{E}[e(G)] = q^{2s} \cdot q^{-1}$. By definition G' contains no copy of $K_{s,c}$, so for $t \geq t_0 := c$, we have shown that there exists a graph G' on at most q^{2s} vertices such that it contains at least $\Omega(q^{2s-1})$ edges and no ncopy of $K_{s,t}$. This gives the desired lower bound when n is a sufficiently large prime prime power. Using Bertrand's postulate gives the desired bound for all n. □

We will admit that on its own Theorem 7.6 is not particularly groundbreaking. Indeed, there exist explicit constructions showing $\mathrm{ex}(n, K_{s,t}) = \Theta(n^{2-1/s})$ when $s > (t-1)!$, and this dependency on t is far better than the implicit constant t_0 one gets from the proof. However, the proof of Theorem 7.6 gives us a new idea for constructing F-free graphs.

And indeed, many developments have been made on this method since Bukh's original construction of this form. In particular, by using more sophisticated tools from algebraic geometry, Bukh [31] showed that Theorem 7.6 holds with $t_0 = C^s$ for some absolute constant C, which stands as the best known bounds for this problem. On the other hand, by carefully modifying the current proof of Theorem 7.6, Conlon [39] was able to construct large graphs avoiding theta graphs $\theta_{a,b}$, which we recall denotes the graph consisting of a internally disjoint paths of length b between two fixed vertices.

Theorem 7.7 ([39]). *For all $b \geq 2$, there exists some $a_0 = a_0(b)$ such that for all $a \geq a_0$, we have*

$$\mathrm{ex}(n, \theta_{a,b}) = \Theta(n^{1+1/b}).$$

Proof Sketch. The upper bound is a result of Faudree and Simonovits [62]. For the lower bound, the key idea is to consider *multiple* random polynomials $f_1, \ldots, f_a \in \mathcal{P}_{d,2b}$ chosen independently, with us defining our graph G on $\mathbb{F}_q^b \cup \mathbb{F}_q^b$ by having $x \sim y$ if and only if $f_i(x, y) = 0$ for all i.

Fix two vertices $x^1, x^{b+1} \in \mathbb{F}_q^b \cup \mathbb{F}_q^b$ and define C to be the tuples of distinct vertices $(x^2, \ldots, x^{b-1})$ such that $x^1 \cdots x^{b+1}$ is a path in G. As before (but with a somewhat more difficult analysis), one can show $\mathbb{E}[|C|^r] = O(1)$. However, in this case C is not quite an algebraic variety because of us requiring C to use distinct vertices. However, one can write C as $X \setminus D$ for two varieties X, D, so Lemma 7.5 still applies and the rest of the proof goes through. □

One can further interpolate between this theorem and Theorem 7.6 to give effective lower bounds on $\mathrm{ex}(n, F)$ whenever F is a "large power of a rooted tree" (e.g. $K_{s,t}$ is just many copies of a star $K_{s,1}$, and $\theta_{a,b}$ is just many copies of a path P_b). This was done by Bukh and Conlon [32] in order to show that for every rational number $r \in [1, 2]$, there exists a finite set of graphs $\mathcal{F}$ such that $\mathrm{ex}(n, \mathcal{F}) = \Theta(n^r)$. The rational exponents conjecture, which says that one can achieve this with $\mathcal{F}$ consisting of a single graph, remains a major open problem.

7.3 Multicolor Ramsey Numbers

Let $r(t; \ell)$ denote the smallest number N such that every ℓ-coloring of $E(K_N)$ contains a monochromatic clique of size t. **Inset history and connection with the earlier Ramsey results proven in the text. Also sketch the proof of the bound you get with the naive method for comparison.**

The following observation will be the key towards going further. The initial idea for this lemma can be seen in Conlon and Ferber [40], though it was first really used by Wigderson [145] and then generalized by Sawin [133].

Lemma 7.8. *Let G be graph with no clique of size t, and let p be the probability that vertices $v_1, \ldots, v_t \in V(G)$ chosen independently and uniformly at random form an independent set. Then for all $\ell \geq 2$, we have*

$$r(t; \ell) \geq p^{-(\ell-2)/t} 2^{(t-1)/2}.$$

Note that when $\ell = 2$ this recovers the usual lower bound for Ramsey numbers from the random coloring.

Proof. Let N be an integer to be determined later, and let $f_1, \ldots, f_{\ell-2} : V(K_N) \to V(G)$ be chosen independently and uniformly at random. Define a coloring $\chi : E(K_N) \to [\ell]$ in the following way: for distinct $x, y \in V(K_n)$, if there exists i such that $f_i(x)f_i(y) \in E(G)$, then set $\chi(xy)$ to be the minimum i with this property. Otherwise, set $\chi(xy)$ to be $\ell - 1$ or ℓ with probability $1/2$ each. That is (as Wigderson notes in his paper), this coloring comes from covering K_N with $\ell - 2$ randomly permuted blowups of G and then randomly using two colors to deal with any uncovered vertices.

We first observe that there is no monochromatic K_t in any color $i \leq \ell-2$. Indeed, if $\{x_1, \ldots, x_t\}$ were such a clique then this would imply $\{f_i(x_1), \ldots, f_i(x_t)\}$ forms a clique in G (since $\chi(x_j x_k) = i$ implies $f_i(x_j)f_i(x_k) \in E(G)$). Thus it remains to show that, with positive probability, there is no monochromatic K_t in color $i \in \{\ell - 1, \ell\}$. Observe that a clique K_t in K_N has all of its edges colored by $\ell - 1$ or ℓ if and only if each f_i maps K_t to an independent set of G, and the probability that this happens is exactly $p^{\ell-2}$ by hypothesis, and from there this K_t will be monochromatic with probability $2^{1-\binom{t}{2}}$. In total then, the expected number of monochromatic cliques will equal $\binom{N}{t} p^{\ell-2} 2^{1-\binom{t}{2}}$, and this will be less than 1 provided $N \leq p^{-(\ell-2)/t} 2^{(t-1)/2}$. Thus there exists a coloring of this size with no monochromatic clique, giving the desired result. □

Observe that the p in Lemma 7.8 roughly corresponds to the number of independent sets of size at most t in G, so we need to find a graph with small clique number and not too many small independent sets. To this end, let $V \subseteq \mathbb{F}_2^t$ be the set of vectors v with $v \cdot v = 0$ (i.e. vectors with even Hamming weight), and let G be the graph on V where two vectors u, v are adjacent if and only if $u \cdot v = 1$.

Lemma 7.9. *If t is even, then the graph G contains no clique of size t.*

Proof. Assume for contradictiont that there exist distinct vectors $v_1, \ldots, v_t \in V$ with $v_i \cdot v_j = 1$ for all $i \neq j$ (and $= 0$ for $i = j$ by definition of V). We claim that these vectors are linearly independent. Indeed, if there exists $\alpha_i \in \{0, 1\}$ with $\sum \alpha_i v_i = 0$, then by taking the dot product of v_j on both sides we find $\sum_{i \neq j} \alpha_i \equiv 0$ for all i, and it is not difficult to show that this implies $\alpha_i = 0$ for all i (here we need that t is even, else $\alpha_i = 1$ for all i would work). However, V is a $t - 1$ dimensional subspace, so it contains no set of t linearly independent vectors, proving the result. □

Lemma 7.10. *The probability p that a uniformly random tuple $(v_1, \ldots, v_t) \in V^t$ is such that $\{v_1, \ldots, v_t\}$ is independent in G is at most $2^{-3t^2/8+o(t^2)}$.*

Proof. Let X be the set of tuples $(v_1, \ldots, v_t) \in V^t$ such that $v_i \cdots v_j = 0$ for all i, j, so our goal is to upper bound $|X|/|V|^t = |X|2^{-t^2}$. Define the rank of a tuple in X to be the rank of the smallest subspaces containing every vertex of the tuple. We claim that the number of tuples in X of rank r is at most

$$t! \left(\prod_{i=0}^{r-1} 2^{t-i} \right) \cdot 2^{(t-r)r} = 2^{tr - \binom{r}{2} + tr - r^2}. \tag{9}$$

Indeed, possibly by reordering the tuple (giving us the factor of $t!$) we can assume the first r vectors are linearly independent, and given $v_1, \ldots, v_i$ with $0 \le i < r$, the number of choices for a v_{i+1} which is linearly independent of $v_1, \ldots, v_i$ is exactly q^{t-i}. After this every vector must lie in the span of $v_1, \ldots, v_r$, giving exactly q^r choices for the remaining $t - r$ vectors.

We next claim that there exists no tuple in X of dimension larger than $t/2$. Indeed, if S is the span of the vectors in a tuple of X, then note that $S \subseteq S^\perp$ since $v_i \cdot v_j = 0$ for all i, j. From linear algebra we have $t = \dim S + \dim S^\perp \ge 2 \dim S$, proving the claim.

It is not hard to prove that (9) is increasing for $r \le t/2$, so plugging in $r = t/2$ gives an upper bound for $|X|/(t/2)$ of the form $2^{5t^2/8 + o(t^2)}$, giving the desired bound on $|X|/|V|^t$. □

Putting all these lemmas together gives the following.

Corollary 7.11. *For $\ell \ge 3$ we have*

$$r(t; \ell) \ge \left(2^{\frac{3\ell}{8} - \frac{1}{4}} \right)^{t - o(t)}.$$

This bound stood as the best for about a year until Sawin [133] realized one could do somewhat better by replacing the algebraic graph G described above with a purely random graph, namely $G_{n,p}$ with $p \approx .455$. Thus, although the initial breakthrough for multicolor Ramsey numbers came from a random algebraic approach, the method was later subsumed by a simpler random model. This sort of thing happens somewhat often with proofs using the random algebraic method. Because of this, some mathematicians are of the opinion that any time the random algebraic method is used, there exists a simpler random model which gives better results. I don't personally believe that this is true, and even if it were, the fact that random algebraic methods consistently give initial breakthroughs to longstanding open problems makes them worth considering in my eyes.

Maybe comment on other ways random homomorphisms are useful.

Part III

Spread Hypergraphs

8 The Spreadness Theorem

Throughout this section we consider hypergraphs $\mathcal{H}$ which may have repeated edges, and we will typically denote the edges of $\mathcal{H}$ by S. We recall that $d(A)$ denotes the degree of a set of vertices A in $\mathcal{H}$, i.e. the number of edges of $\mathcal{H}$ containing A.

We say that a hypergraph $\mathcal{H}$ is r-bounded if all of its edges have size at most r. We say that a hypergraph $\mathcal{H}$ is q-spread[1] for some $0 < q < 1$ if $\mathcal{H}$ is non-empty and if $d(A) \leq q^{|A|}|\mathcal{H}|$ for all sets of vertices A. The main result for q-spread hypergraphs is the following.

Theorem 8.1 ([9, 68])**.** *Let $\mathcal{H}$ be an r-bounded q-spread hypergraph on V. There exists an absolute constant K_0 such that if W is a set of size $Cq \log r \cdot |V|$ chosen uniformly at random from V with $C \geq K_0$, then*

$$\Pr[W \text{ contains an edge of } \mathcal{H}] \geq 1 - 8C^{-1}.$$

We note that better quantitative versions of Theorem 8.1 exist, see e.g. **Tao's reformulation below?**, but as stated this theorem already does a lot. Let's start by looking at some applications before turning it's short (though very dense!) proof.

8.1 Applications

Our first application is the following.

Theorem 8.2. *Let $G^r_{n,m}$ be the r-graph chosen uniformly at random amongst all r-graphs with n vertices and m edges. Then there exists a constant C such that if $m \geq Cn \log n$ and n is a multiple of r, then $G^r_{n,m}$ contains a perfect matching a.a.s.* [2]

It is not too difficult to show that this bound on m is essentially best possible. We note that morally speaking, $G^2_{n,m}$ acts the same as $G_{n,p}$ where $p = m/\binom{n}{2}$. In particular, one can use Theorem 8.2 to prove that $G_{n,p}$ contains a perfect matching a.a.s. if $p = \Omega(\log n/n)$. Proving Theorem 8.2 for $r = 2$ is not hard, but the result for general r was thought to be very difficult, with its first proof due to Johansson, Kahn, and Vu [95] using a rather involved argument. We will prove Theorem 8.2 in just a few lines with Theorem 8.1.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K^r_n)$ where each hyperedge S is a perfect matching of

[1]Some texts would say that such an $\mathcal{H}$ is q^{-1}-spread.

[2]This means "asymptotically almost surely", i.e. the probability of this event happening tends to 1 as n tends to infinity.

K_n^r. Observe that for any set $A \subseteq E(K_n^r)$, we have

$$\begin{aligned} d(A) \cdot |\mathcal{H}|^{-1} &= \frac{(n-r|A|)!}{(r!)^{n/r-|A|}(n/r-|A|)!} \cdot \frac{(r!)^{n/r}(n/r)!}{n!} \\ &= (r!)^{|A|}\binom{n/r}{|A|}\binom{n}{r|A|}^{-1}\frac{|A|!}{(r|A|)!} \\ &\le (r!)^{|A|}(en/r|A|)^{|A|} \cdot (n/r|A|)^{-r|A|} \cdot (|A|)^{|A|} \cdot (r|A|/e)^{-r|A|} \\ &= (r!)^{|A|}e^{(r+1)|A|}n^{-(r-1)|A|} \le (n/re^3)^{-(r-1)|A|}. \end{aligned}$$

Thus $\mathcal{H}$ is $(n/re^3)^{-r+1}$-spread. It is also (n/r)-uniform and has a ground set $V = E(K_n^r)$ of size $\binom{n}{r}$. By Theorem 8.1, we see that if m is at least as large as in our hypothesis, then with high probability a random m-subset of $\mathcal{H}$ will contain a hyperedge, i.e. $H_{n,m}^r$ will contain a perfect matching with high probability. □

Another basic example is the following.

Proposition 8.3. *Let F be an r-graph and define $t(F) = \max\{|E(F')|/|V(F')| : F' \subseteq F\}$. Let $G_{n,m}^r$ be as in Theorem 8.2. There exists a constant $C(F)$ such that if $m \ge C(F)n^{r-1/t(F)}$, then $G_{n,m}^r$ contains a copy of F a.a.s.*

A simple first moment argument shows that this bound is tight. One can prove Proposition 8.3 using a standard but somewhat tedious second moment argument, but using Theorem 8.1 gives a shorter proof.

Proof. Let $\mathcal{H}$ be the hypergraph on $E(K_n^r)$ whose hyperedges correspond to copies of F. Observe that $\mathcal{H}$ being q-spread is equivalent to having $(d(A)/|\mathcal{H}|)^{1/|A|} \le q$ for all $A \subseteq V = E(K_n^r)$. Any set $A \subseteq E(K_n^r)$ of positive degree in $\mathcal{H}$ forms a subgraph $F' \subseteq F$ with $|E(F')| = |A|$, and in this case

$$\left(\frac{d(A)}{|\mathcal{H}|}\right)^{1/|A|} \le \left(\frac{n^{|V(F)|-|V(F')|}}{\binom{n}{|V(F)|}}\right)^{1/|A|} \le |V(F)|^{|V(F)|} \cdot n^{-|V(F')|/|E(F')|}.$$

Thus we see that $\mathcal{H}$ is q-spread with

$$q = \max\{|V(F)|^{|V(F)|} \cdot n^{-|V(F')|/|E(F')|} : F' \subseteq F\} = |V(F)|^{|V(F)|} \cdot n^{-1/t(F)}.$$

Plugging this into Theorem 8.1 gives the result. □

The study of q-spread hypergraphs was initiated by Alweiss, Lovett, Wu, and Zhang [9] where they proved a slightly weaker version of Theorem 8.1. Their motivation came from the Erdős sunflower conjecture. A *k-sunflower* is a hypergraph with edges $S_1, \ldots, S_k$ such that there exists a set K called the *kernel* which has $S_i \cap S_j = K$ for all $i \ne j$.

Theorem 8.4. *There exists an absolute constant $C > 0$ such that if $\mathcal{H}$ is an r-graph with at least $(Ck\log r)^r$ edges, then $\mathcal{H}$ contains a k-sunflower.*

We note that [9] was the first to prove bounds of the form $(\log r)^{r+o(1)}$ for fixed k, with [124, 19] later giving better bounds in terms of k. Prior to [9], the best known bounds were of the form $r^{r-o(1)}$. It is a famous conjecture of Erdős that one can prove a bound of the form $c_k^{r+o(1)}$.

Proof. We prove the result by induction on r, the $r = 1$ case being trivial. Let $\mathcal{H}$ be an r-graph with at least $(Ck \log r)^r$ edges. If $\mathcal{H}$ is not q-spread with $q = (Ck \log r)^{-1}$, then there exists some $A \subseteq V(H)$ such that $d(A) \geq (Ck \log r)^{r-|A|}$. This means that the link hypergraph $\mathcal{H}_A = \{S \setminus A : S \in \mathcal{H}, A \subseteq S\}$ has size at least $(Ck \log r)^{r-|A|}$. Since $\mathcal{H}_A$ is an $(r - |A|)$-uniform hypergraph, by induction $\mathcal{H}_A$ contains a k-sunflower, say with edges $S_1 \setminus A, \ldots, S_k \setminus A \in \mathcal{H}_A$. It is not difficult to check that $S_1, \ldots, S_k \in \mathcal{H}$ forms a k-sunflower in $\mathcal{H}$. We conclude that any $\mathcal{H}$ with at least $(Ck \log r)^r$ edges which is not q-spread contains a k-sunflower, so from now on we may assume $\mathcal{H}$ is q-spread.

Possibly by adding isolated vertices to $\mathcal{H}$, we can assume that the size of the vertex set V of $\mathcal{H}$ is a multiple of $2k$. Let $V_1, \ldots, V_{2k}$ be a random partition of V such that each $V_i \subseteq V$ has size $(2k)^{-1}|V|$. This means that each V_i is a uniformly chosen set of V of size $(2k)^{-1}|V| = \frac{1}{2}C(\log r)q|V|$. Let 1_i be the indicator variable for the event that V_i contains an edge of $\mathcal{H}$. By Theorem 8.1, we have $\Pr[1_i = 1] \geq \frac{1}{2}$ provided C is sufficiently large. In this case, $\mathbb{E}[\sum 1_i] \geq k$, and hence there exists some partition $V_1, \ldots, V_{2k}$ such that $\sum 1_i \geq k$, which in particular means there exist k disjoint edges of $\mathcal{H}$. This is a k-sunflower in $\mathcal{H}$, proving the result. □

8.2 Proof of the Spreadness Theorem

A maybe simpler proof of this/Park-Pham can be found in Rao's talk on "The Sunflower Lemma and Monotone Thresholds", see also the link commented out below

There are by now a number of proofs of Theorem 8.1, though most of them maintain the same core set of ideas. The proof we present here is based off of a proof due to Rao [125] which gives weaker quantitative bounds. We emphasize that while the proof itself is very short, it is also very dense in content, so we'll spend some time trying to build up some intuition for it.

Recall that $\mathcal{H}$ is an r-bounded q-spread hypergraph on V, and that we want to show that a uniformly random set W of size $Cq \log r \cdot |V|$ contains an edge of $\mathcal{H}$ with high probability. In order to use an iterative approach, we consider a uniform random vector of disjoint sets $(W_1, \ldots, W_{\log r})$ each of size $Cq|V|$. Let $W_{\leq i} = \bigcup_{j \leq i} W_j$, and note that W has the same distribution as $W_{\log r}$, so it suffices to work with these random sets.

A super ideal situation for our iterative approach would be if for all $S \in \mathcal{H}$, we have $|S - W_{\leq i}| < 2^{-i}r$. Indeed, with this at $i = \log r$, we would get that every edge is contained in $W_{\leq \log r}$. Of course, this is far too much to hope for. However, since we only need $W_{\leq \log r}$ to contain a single edge, it would suffice to have this work out for *some* S. As such it perhaps make sense to say that an edge S "succeeds" at step i if $|S - W_{\leq i}| < 2^{-i}r$, and then to argue that with high probability some edge succeeds at each step. Unfortunately this notion of success is too restrictive to work. The key insight is that we can loosen our condition by saying that an edge S "succeeds" if there exists some edge $S' \subseteq S \cup W_{\leq i}$ (or equivalently $S' - W_{\leq i} \subseteq S - W_{\leq i}$) such that $|S' - W_{\leq i}| < 2^{-i}r$. The point is that (1) this condition is easier to achieve (and in particular will be achieved by "most" S at each step), and (2) if some S succeeds at each step, then the S' it points to for step $\log r$ will be contained in $W_{\leq \log r}$. Need to fact check that this is really what's going on, and in particular we maybe need S' also to have not failed at this point.

With this in mind, given $W_1, \ldots, W_i$, we iteratively define "failure" hypergraph $\mathcal{F}_i$ to be those

$S \notin \mathcal{F}_{\le i-1} := \bigcup_{j \le i-1} \mathcal{F}_j$ (i.e. which haven't failed at any previous step) such that for all $S' \in \mathcal{H} - \mathcal{F}_{\le i-1}$ satisfying $S' \subseteq S \cup W_{\le i}$, we have $|S' - W_{\le i}| \ge 2^{-i} r$. The key claim is the following.

Lemma 8.5. *Given $W_{\le i-1}$, we have $\mathbb{E}[|\mathcal{F}_i|] \le 2(C/4)^{-2^{-i}r}|\mathcal{H}|$.*

To upper bound $|\mathcal{F}_i|$, it will help to instead upper bound the size of an auxiliary hypergraph defined as follows. Given $W_1, \ldots, W_i$, we define the *fragment* $T(S, W_{\le i})$ of an edge $S \in \mathcal{H} - \mathcal{F}_{i-1}$ to be a set of minimum size in $\{S' - W_{\le i} : S' \in \mathcal{H} - \mathcal{F}_{\le i-1}, S' \subseteq S \cup W_{\le i}\}$ (say the lexicographically smallest set if there are multiple of minimum size). We let $\mathcal{G}_i$ be the hypergraph where T is an edge if $T = T(S, W_{\le i})$ for some $S \in \mathcal{F}_i$. Note that by definition this means $|T| \ge 2^{-i} r$, and that for every $S \in \mathcal{F}_i$, there exists some $T \in \mathcal{G}_i$ with $T \subseteq S$. This last condition says $\mathcal{G}_i$ is an *undercover* of $\mathcal{F}_i$, which will also be a key condition in our upcoming proof of the Park-Pham Theorem.

Proof. Let $w := Cq|V|$, which we recall is the size of W_i, and let $n_i = |V - W_{\le i-1}|$. Let $\mathcal{P}$ consist of all pairs (S, W) with $S \in \mathcal{H}$ and $W \in \binom{V - W_{\le i-1}}{w}$ such that $S \in \mathcal{F}_i$ whenever $W_i = W$. Similarly given an integer $a \ge 2^{-i} r$, let $\mathcal{P}_a$ consist of all pairs (T, W) with $|T| = a$ and $W \in \binom{V - W_{\le i-1}}{w}$ such that $T \in \mathcal{G}_i$ whenever $W_i = W$. We claim that. **Probably use t instead of a**

$$\mathbb{E}[|\mathcal{F}_i|] = |\mathcal{P}|\binom{n_i}{w}^{-1} \le \sum_{a \ge 2^{-i} r} q^a |\mathcal{H}||\mathcal{P}_a|\binom{n_i}{w}^{-1}. \tag{10}$$

Indeed, the equality is straightforward. Because $\mathcal{G}_i$ is an undercover of $\mathcal{F}_i$, for every pair $(S, W) \in \mathcal{P}$ there exists a pair $(T, W) \in \bigcup_a \mathcal{P}_a$ such that $T \subseteq S$. Moreover, for each set T of size a, the number of $S \in \mathcal{H}$ with $T \subseteq A$ is at most $q^a|\mathcal{H}|$ by the definition of $\mathcal{H}$ being q-spread. This gives the stated inequality

It remains to count the number of elements $(T, W) \in \mathcal{P}_a$. We will identify such a pair by first specifying the set $T \cup W$, and then specifying T (which uniquely determines W). We first note that $T \cup W$ is a set of size $a + w$, so the number of choices for this step is at most

$$\binom{n_i}{a + w} \le (n_i/w)^a \cdot \binom{n_i}{w} = (Cq)^{-a}\binom{n_i}{w}.$$

Given $T \cup W$, choose any $S' \in \mathcal{H} - \mathcal{F}_{\le i-1}$ with $S' - W_{\le i-1} \subseteq T \cup W$. Crucially, we must have $T \subseteq S' - W_{\le i-1}$, as otherwise if $T = T(S, W_{\le i-1} \cup W)$ for some S, then taking $T' = S' - (W_{\le i-1} \cup W) \subsetneq T$ (with the inclusion holding because $S' - W_{\le i-1} \subseteq T \cup W$, and the strictness holding if $T \not\subseteq S' - W_{\le i-1}$), we find that T cannot be the fragment of S (since T' is a smaller set than T satisfying the same properties). Note that $S' \notin \mathcal{F}_{\le i-1}$ implies $|S' - W_{\le i-1}| \le 2^{-i+1} r$, so the number of choices for T is at most $2^{2^{-i+1} r} = 4^{2^{-i} r}$.

In total we conclude that $|\mathcal{P}_a| \le (Cq)^{-a} 4^{2^{-i} r}\binom{n_i}{w}$. Plugging this into (10), we find

$$\mathbb{E}[|\mathcal{F}_i|] \le \sum_{a \ge 2^{-i} r} q^a |\mathcal{H}| \cdot (Cq)^{-a} 4^{2^{-i} r} = 4^{2^{-i} r}|\mathcal{H}| \sum_{a \ge 2^{-i} r} C^{-a} \le 2(C/4)^{-2^{-i} r}|\mathcal{H}|,$$

with this last step holding for C sufficiently small. $\square$

With this lemma, we have

$$\mathbb{E}[|\mathcal{F}_{\le \log r}|] \le \sum_{i=1}^{r} 2(C/4)^{-2^{-i}r}|\mathcal{H}| \le 16C|\mathcal{H}|.$$

By Markov, the probability that $|\mathcal{F}_{\le \log r}| = |\mathcal{H}|$ is at most $\frac{1}{16C}$, so $\mathcal{F}_{\le \log r} \neq \mathcal{H}$ is at least $1-\frac{1}{16C}$. As noted above, if there exists $S \in \mathcal{H} - \mathcal{F}_{\le \log r}$ then S points to an edge which is contained in $W_{\le \log r}$, so we conclude that this random set $W_{\le \log r}$ of size $Cq\log(r)|V|$ contains an edge with probability at least $1-\frac{1}{16C}$.

8.3 Losing Logarithms

As we noted earlier, the bound of Theorem 8.2 is best possible. In particular, the $\log r$ term of Theorem 8.1 is necessary in general. However, under certain conditions one can remove this logarithmic term. This was first observed by Kahn, Narayanan, and Park [97] where they found tight bounds on the threshold of a square of a Hamiltonian cycle in $G_{n,p}$. Here we briefly outline how, under special circumstances, one can modify the previous proof to get rid of the $\log r$ factor.

The main idea is that instead of setting our cutoff points for our fragments to be $r/2, r/4, \ldots$, we instead set them to be $k_1, k_2, \ldots$ for some suitable sequence k_i with significantly fewer than $\log r$ terms. In this setup, one could try to naively go through Lemma 8.5 and replace $2^{-i+1}r$ with k_{i-1} and $2^{-i}r$ with k_i, which will roughly give us

$$\mathbb{E}[|\mathcal{F}_i|] \le 2^{k_{i-1}} C^{-k_i} |\mathcal{H}|,$$

but this will be terrible unless k_{i-1} differs from k_i by a multiplicative constant depending on C.

One way we can get around this is if we impose that for all sets A and integers j with $k_{i-1} \ge |A| \ge j \ge k_i$, we have that the number of edges S' with $|A \cap S'| = j$ is at most $q^j|\mathcal{H}|$. Note that this is stronger than spreadness since, when taking $j = |A|$, the condition $|A \cap S'| = j$ just says S' is an edge containing A, so this bound exactly says $\deg_{\mathcal{H}}(A) \le q^{|A|}|\mathcal{H}|$, which is the spreadness condition. Assuming this condition holds, we will count the pairs $(S, W) \in \mathcal{P}$ in a more subtle way.

Let $T = T(S, W_{\le i-1} \cup W)$, which we note has $|T| \ge k_i$ and $T \subseteq S$. We first specify $T \cup W$, which as before can be done in roughly $(Cq)^{-a}\binom{n_i}{w}$ ways. We then pick some edge $S' \notin \mathcal{F}_{\le i-1}$ such that $A := S' - W_{\le i-1} \subseteq T \cup W$, where as before we have $T \subseteq A$. Since $T \subseteq A \cap S$, we have $j := |A \cap S| \ge a$. Given j, the number of S with $|A \cap S| = j$ is at most $q^j|\mathcal{H}|$ by our condition (and we have $|A| \le k_{i-1}$ as otherwise we would have $S' \in \mathcal{F}_{\le i-1}$). Since we now know S and $T \cup W$, we also know $S \cup W$ (since $T \subseteq S$), and hence T (since T is purely a function of the set $S \cup W$ and $W_{\le i-1}$ Need to double check this; in any case there are trivially at most 2^j choices for $T \subseteq A \cap S$), and hence $W = (T \cup W) \setminus T$ (since T is disjoint from W by definition). With this, we see that the total number of choices is at most

$$\sum_{a \ge k_i} C^{-a}\binom{n_i}{w} \sum_{j \ge a} q^j |\mathcal{H}| \approx C^{-k_i}\binom{n_i}{w}|\mathcal{H}|,$$

giving the desired result. A more formal theorem/proof can be found in [139], though the approach used there is an older and more complicated version of the one presented here. See also [71] for a proof in the specific case of getting rid of the logarithm for the square of a Hamiltonian cycle.

9 TODOThe Park-Pham Theorem

10 Spread Approximations

Motivated by our success in proving bounds on hypergraphs which don't contain sunflowers in Theorem 8.4, we consider some additional applications of spread hypergrpahs to extremal set theory. For this it will be convenient to define a slightly different notion of spreadness which was introduced by Kupavskii and Zakharov [106].

Given a set of vertices A, we define the link hypergraph $\mathcal{H}(A) = \{e \setminus A : e \in \mathcal{H},\ A \subseteq e\}$. We say that an n-vertex r-graph $\mathcal{H}$ is τ-*homogeneous* if

$$d(A) = |\mathcal{H}(A)| \leq \tau^{|A|} \frac{\binom{n-|A|}{r-|A|}}{\binom{n}{r}} |\mathcal{H}|.$$

Intuitively, τ-homogeneous hypergraphs should be thought of as (τ/n)-spread hypergraphs. Formally, we have the following.

Lemma 10.1. *If $\mathcal{H}$ is an n-vertex r-graph which is τ-homogeneous, then it is q-spread with $q = \frac{\tau r}{n}$.*

Proof. Note that $\binom{n-1}{r-1} = \frac{r}{n}\binom{n}{r}$. By repeating this logic and using that $\frac{r-i}{n-i} \leq \frac{r}{n}$ whenever $i \leq r \leq n$, we find

$$d(A) \leq \tau^{|A|} \frac{\binom{n-|A|}{r-|A|}}{\binom{n}{r}} |\mathcal{H}| \leq \left(\frac{\tau r}{n}\right)^{|A|} |\mathcal{H}|,$$

proving the result. $\square$

The main motivation for this definition is the following.

Theorem 10.2 ([106])**.** *Fix $k \in \mathbb{Z}_{\geq 1}$ and $\tau \in \mathbb{R}_{\geq 1}$. For every n-vertex r-graph $\mathcal{H}$, there exists an "approximation" hypergraph $\mathcal{S}$ with edges of size at most k, and a "remainder" $\mathcal{H}' \subseteq \mathcal{H}$ with the following properties:*

- *For every $e \in \mathcal{H} \setminus \mathcal{H}'$, there exists $S \in \mathcal{S}$ such that $S \subseteq e$.*
- $|\mathcal{H}'| \leq \tau^{-k-1}\binom{n}{r}$.
- *For every $S \in \mathcal{S}$ there exists $\mathcal{H}_S \subseteq \mathcal{H}$ such that the link hypergraph $\mathcal{H}_S(S)$ is τ-homogeneous.*

This first condition says that $\mathcal{S}$ "approximates" $\mathcal{H} \setminus \mathcal{H}'$ in the sense that it's an undercover, while the second condition says that "most" of $\mathcal{H}$ is approximated. The third condition gives us the very useful property that each edge S of our approximating hypergraph $\mathcal{S}$ has a well spread link, which will imply that a random partition of our vertex set will give a sunflower whose kernel contains S.

Add more intuition for this proof, possibly via recalling the proof of the improved sunflower bounds (namely you go by picking a largest S such that $\mathcal{F}(S)$ isn't spread).

Proof. If $n \le k$, then we can take $\mathcal{S} = \mathcal{H}$, $\mathcal{H}_S = \{S\}$, and $\mathcal{H}' = \emptyset$ to get the result, so we may assume $n > k$. For this proof we define $\mathcal{H}[S] = \{e : e \in \mathcal{H},\ S \subseteq e\}$, i.e. this is just the link hypergraph $\mathcal{H}(S)$ after adding the set S back into each edge.

Let $\mathcal{H}^1 = \mathcal{H}$. Given $\mathcal{H}^i$, let S_i be a maximal set of vertices with $d_{\mathcal{H}^i}(S_i) \ge \tau^{|S_i|} \frac{\binom{n-|S_i|}{r-|S_i|}}{\binom{n}{r}} |\mathcal{H}^i|$. If $|S_i| > k$ (which happens automatically if $\mathcal{H}^i = \emptyset$ and $n > k$), then stop the procedure. Otherwise set $\mathcal{H}^{i+1} = \mathcal{H}^i \setminus \mathcal{H}^i[S_i]$.

Say we stop this procedure at $i = m+1$. Set $\mathcal{H}' = \mathcal{H}^{m+1}$, $\mathcal{S} = \{S_i : i \le m\}$, and $\mathcal{H}_{S_i} = \mathcal{H}^i[S_i]$. Observe that by construction each edge of $\mathcal{S}$ has size at most k and that the first condition of the theorem is satisfied (since $e \in \mathcal{H} \setminus \mathcal{H}'$ implies $e \in \mathcal{H}^i$ for some $i \le m$, and taking $S = S_i$ works). Again by construction, we have

$$|\mathcal{H}'| = |\mathcal{H}^{m+1}| \le d_{\mathcal{H}^{m+1}}(S_{m+1}) \tau^{-|S_i|} \frac{\binom{n}{r}}{\binom{n-|S_{m+1}|}{r-|S_{m+1}|}} \le \tau^{-k-1} \binom{n}{r},$$

where this last step used the trivial bound $d_{\mathcal{H}^{m+1}}(S_{m+1}) \le \binom{n-|S_{m+1}|}{r-|S_{m+1}|}$ and that $|S_{m+1}| \ge k+1$ by assumption of us stopping the procedure here. This establishes the second condition.

For the third condition, note that for any set T disjoint from S_i, we have

$$d_{\mathcal{H}_{S_i}}(T) = d_{\mathcal{H}^i}(S \cup T) < \tau^{|S_i \cup T|} \frac{\binom{n-|S_i \cup T|}{r-|S_i \cup T|}}{\binom{n}{r}} |\mathcal{H}^i|,$$

where this last step used the maximality of S_i. We also have

$$|\mathcal{H}^i| \le \tau^{-|S_i|} \frac{\binom{n}{r}}{\binom{n-|S_i|}{r-|S_i|}} d_{\mathcal{H}^i}(S_i) = \tau^{-|S_i|} \frac{\binom{n}{r}}{\binom{n-|S_i|}{r-|S_i|}} |\mathcal{H}_{S_i}|.$$

Combining these two inequalities gives the third condition, proving the result. $\square$

It turns out that when τ is small, the $\mathcal{S}$ given by Theorem 10.2 inherits intersection properties of the original hypergraph $\mathcal{H}$. Recall that a hypergraph is t-intersecting if every two edges intersect in at least t vertices.

Lemma 10.3. *There exists an absolute constant C such that the following holds. Let $\mathcal{H}$ be an n-vertex r-graph, and let $\mathcal{S}$ by the hypergraph guaranteed by Theorem 10.2 with parameters k, τ. If $n \ge C\tau r \max\{\log r, k\}$, then*

$$\{S \cap T : S, T \in \mathcal{S}\} \subseteq \{e \cap f : e, f \in \mathcal{H}\}.$$

We emphasize that we allow $S = T$ in the lemma statement.

Proof. Let $S, T \in \mathcal{S}$, and note that Lemma 10.1 implies $\mathcal{H}_S(S)$ is $\frac{\tau r}{n}$-spread. Let $\mathcal{H}'_S \subseteq \mathcal{H}_S(S)$ be the hypergraph obtained by deleting all of the vertices $x \in T \setminus S$. Then

$$|\mathcal{H}'_S| \ge |\mathcal{H}_S(S)| - \sum_{x \in T \setminus S} d_{\mathcal{H}_S(S)}(x) \ge (1 - \frac{\tau r}{n} \cdot k)|\mathcal{H}_S(S)| \ge \frac{1}{2}|\mathcal{H}_S(S)|,$$

where the second inequality used that $\mathcal{H}_S(S)$ is spread, and the last inequality holds for $C \geq 2$. This together with $\mathcal{H}_S(S)$ being $\frac{\tau r}{n}$-spread implies $\mathcal{H}'_S$ is $q := \frac{2\tau r}{n}$-spread. Similarly if one defines $\mathcal{H}'_T \subseteq \mathcal{H}_T(T)$ by deleting the vertices of $S \setminus T$ we get that this is q-spread.

Randomly partition the vertices of $V(\mathcal{H}) \setminus (S \cup T)$ into two sets V_1, V_2 of size at least $\frac{1}{2}(n-2k) \geq \frac{1}{4}n$, where this holds if C is sufficiently large. By hypothesis this is at least $\frac{1}{4}C\tau r \log r = \frac{1}{8}Cq \log rn$. Thus by Theorem 8.1, if C is sufficiently large, then with positive probability both V_1, V_2 contain edges e', f' of $\mathcal{H}'_S, \mathcal{H}'_T$ respectively, which by definition means e', f' contains no vertices of T, S. This means $e = e' \cup S,\ f = f' \cup T$ are edges of $\mathcal{H}$ with $e \cap f = S \cap T$, proving the result. $\square$

A simple application of this result gives the t-intersecting version of the Erdős-Ko-Rado theorem (albeit with suboptimal bounds on n).

Theorem 10.4. *Let $\mathcal{H}$ be an n-vertex r-graph such that $|e \cap f| \geq t$ for all $e, f \in \mathcal{H}$. If n is sufficiently large in terms of r, then $|\mathcal{H}| \leq \binom{n-t}{r-t}$ with equality holding if and only if $\mathcal{H}$ consists of every edge containing some fixed set T of size t.*

Proof. Apply Theorem 10.2 with $k = t$ and $\tau = \frac{n}{Cr \max\{\log r, k\}}$ with C as in Lemma 10.3, and let $\mathcal{H}', \mathcal{S}$ be the resulting families. Note that by Theorem 10.2,

$$|\mathcal{H}'| \leq \tau^{-k-1}\binom{n}{r} = O(n^{r-t-1}).$$

By Lemma 10.3, our hypothesis on $\mathcal{H}$, and the fact that $|S| \leq k = t$ for all $S \in \mathcal{S}$, we see that $\mathcal{S}$ is either empty or consists of a single set T of size t. In the former case $|\mathcal{H}| = |\mathcal{H}'| = O(n^{r-t-1}) < \binom{n-t}{r-t}$ and there is nothing to prove, so we may assume such a T exists. This implies that every element of $\mathcal{H} \setminus \mathcal{H}'$ contains T.

First consider the case that $\mathcal{H}$ contains some e which does not contain T. By our observation above, this means that every element of $\mathcal{H} \setminus \mathcal{H}'$ contains both T and some additional element of e. This implies

$$|\mathcal{H} \setminus \mathcal{H}'| + |\mathcal{H}'| \leq r\binom{n-t-1}{r-t-1} + |\mathcal{H}'| = O(n^{r-t-1}) < \binom{n-t}{r-t}.$$

Thus we can assume every element of $\mathcal{H}$ contains T, which means $|\mathcal{H}| \leq \binom{n-t}{r-t}$. $\square$

The above argument actually gives the following stability result: for all r, t there exists a constant $c' = c'(r,t)$ such that if $\mathcal{H}$ is t-intersecting with $|\mathcal{H}| > c'\binom{n-t}{r-t}$, then there exists a set of size t which is contained in every edge of $\mathcal{H}$.

Another application is a bound for how large an intersecting hypergraph $\mathcal{H}$ can be if it's "far" from the extremal example, i.e. a star. There are many ways to make the notion of "far" precise. One way is to demand that every vertex of $\mathcal{H}$ to be contained in the same number of edges, i.e. to demand that $\mathcal{H}$ be regular.

Theorem 10.5 ([106])**.** *There exists an absolute constant $C > 0$ such that if $n \geq Cr \max\{\log r, k\}$, then any n-vertex intersecting r-graph $\mathcal{H}$ has $|\mathcal{H}| \leq 2^{-k}\binom{n}{r}$.*

This result is roughly optimized when $k \approx n/r$, giving an upper bound of roughly $2^{-n/r}\binom{n}{r}$, which is typically much stronger than the bound $\binom{n-1}{r-1}$ given by Erdős-Ko-Rado for (not necessarily regular) intersecting families. Note that in particular this bound implies that regular intersecting hypergraphs cannot exist if $n \gg r$ **Which is maybe obvious by elementary means.**

Proof. Apply Theorem 10.2 with $\tau = 2$, which we can do if C is sufficiently large. Let $\mathcal{H}', \mathcal{S}$ be the corresponding families that we get. Note that by Theorem 10.2 we have

$$|\mathcal{H} \setminus \mathcal{H}'| \geq \frac{1}{2}|\mathcal{H}|.$$

We claim that every edge $e \in \mathcal{H} \setminus \mathcal{H}'$ intersects every $S \in \mathcal{S}$. Indeed, by the first property of Theorem 10.2, there exists some $S' \subseteq e$ with $S' \in \mathcal{S}$, and by Lemma 10.3, this S' (and hence e) intersects S.

Pick any $S \in \mathcal{S}$. By the previous claim, some vertex $x \in S$ must satisfy

$$d(x) \geq \frac{1}{|S|}|\mathcal{H} \setminus \mathcal{H}'| \geq \frac{1}{2k}|\mathcal{H}|,$$

where this last step used $|S| \leq k$ for any $S \in \mathcal{S}$ and $|\mathcal{H} \setminus \mathcal{H}'| \geq \frac{1}{2}|\mathcal{H}|$. However, since $\mathcal{H}$ is regular, by the handshaking lemma we must have

$$d(x) = \frac{r}{n}|\mathcal{H}|.$$

This contradicts the previous bound if $n > 2rk$, giving the result. □

10.1 Further Results

One can push the ideas of this section significantly further. Here we sketch out some of these ideas, and we refer the reader to [106] for the details.

One important direction is that one can consider different "ambient families." That is, up to this point we were considering r-uniform hypergraphs, i.e. $\mathcal{H} \subseteq \mathcal{A} := \binom{[n]}{r}$. Alternatively, one can identify subsets of $[n^2]$ by their 0-1 characteristic vectors, which can in turn be written as n-dimensional 0-1 matrices. In particular, by letting $\mathcal{A}$ denote the set of such vectors corresponding to permutation matrices, we can now consider "intersection" problems for sets of permutations $\mathcal{H} \subseteq \mathcal{A}$. Here one can again go through similar steps to develop a notion of homongenous families (with respect to our new choice of $\mathcal{A}$) in order to get results about sets of permutations $\mathcal{H}$ which are intersecting (i.e. such that any two permutations $\pi, \sigma \in \mathcal{H}$ have $\pi(i) = \sigma(i)$ for some i, i.e. if there corresponding 0-1 matrices have a common entry equal to 1).

Through Lemma 10.3, we showed that if $\mathcal{H}$ is t-intersecting and τ is small enough, then we can guarantee that $\mathcal{S}$ is t-intersecting. Through a more refined argument, one can show that if τ is very small, then in fact this same conclusion holds if we only impose the much weaker condition that $|e \cap f| \neq t - 1$ for $e, f \in \mathcal{H}$.

Part IV

Entropy

11 Introduction

Throughout this part we let log denote logarithms base 2 unless stated otherwise, and we define $x \log x = 0$ whenever $x = 0$.

Let X be a discrete random variable and $p_x = \Pr[X = x]$. The *binary entropy* of this random variable is defined as

$$H[X] = -\sum_{x \in \mathrm{supp}(X)} p_x \log(p_x),$$

where the sum ranges over all x in the support of X (i.e. those x with $p_x > 0$).

The definition for $H[X]$ is quite strange if one has never seen it before. Roughly speaking, $H[X]$ can be thought of as measuring how much "information" the random variable X carries. For example, one can easily check that $H[X] = 0$ if and only if X is deterministic, corresponding to the fact that knowing the outcome of a deterministic process gives no new information.

11.1 The Main Properties

Before getting into applications, let us start by recording the most common properties of entropy that will be used throughout this part, and for this it will be useful to establish some notation.

Given a random variable X and an event E, we define $X|E$ to be the random variable X conditioned on the event E. Given a pair of random variables X, Y we define

$$H[X|Y] := \mathbb{E}_{y \sim Y} H[X|Y = y] = -\sum_y \Pr[Y = y] \sum_x \Pr[X = x|Y = y] \log(\Pr[X = x|Y = y]).$$

The expression $H[X|Y]$ is commonly referred to as *conditional entropy.* For convenience, we will often denote vectors of random variables $(X_1, \ldots, X_n)$ as simply $X_1, \ldots, X_n$, e.g. by writing $H[X_1, \ldots, X_n]$ instead of $H[(X_1, \ldots, X_n)]$. For an integer i we let $X_{<i} := (X_1, \ldots, X_{i-1})$. Slightly more generally, if X is a random vector indexed by a set S with a total ordering $<$, then we let $X_{<s}$ denote the elements of X indexed by $t < s$.

We now state our list of properties about the entropy function. The reader is not expected to memorize this right away, though it might be a good idea to what extent these properties agree with the intuition[1] that $H[X]$ measures the information of X.

Claim 11.1. *The following properties hold for any random variable X.*

[1]For example, Subadditivity says that the total information in the vector (X, Y) is at most the sum of the information of X and Y, Dropping Conditioning says that knowing less at the start can only lead to more information gained, and Data Processing says that one can't do anything to a random variable Y in order to give it more information than it already has.

- *(Non-negativity) We have* $H[X] \geq 0$ *with equality if and only if* X *is deterministic.*

- *(Maximality Principle) We have*

$$H[X] \leq \log |\mathrm{supp}(X)|,$$

with equality if and only if X *is uniformly distributed on* $\mathrm{supp}(X)$*.*

- *(Chain Rule) For two random variables* X, Y *we have*

$$H[X, Y] = H[X] + H[Y|X].$$

More generally, given random variables $X_1, \ldots, X_n$ *we have*

$$H[X_1, \ldots, X_n] = \sum_i H[X_i | X_1, \ldots, X_{i-1}].$$

- *(Subadditivity) For random variables* $X_1, \ldots, X_n$ *we have*

$$H[X_1, \ldots, X_n] \leq \sum_i H[X_i].$$

- *(Dropping Conditioning) For random variables* X, Y, Z *we have*

$$H[X|Y] \leq H[X],$$

with equality if and only if X *is independent of* Y*. Similarly*

$$H[X|Y, Z] \leq H[X|Y],$$

with equality if and only if X *conditioned on* Y *has the same distribution as* X *conditioned on both* Y *and* Z*.*

- *(Data Processing Inequality) If* X, Y, Z *are random variables such that* Z *is a function of* Y*, then*

$$H[X|Z] \leq H[X|Y].$$

Not only does the function $H[X] = -\sum_x p_x \log(p_x)$ satisfy these properties, it is in fact essentially the unique function satisfying (somewhat weaker versions of) these properties. Because of this, H will be the "right" function for us to use to encode how much "information" X has. For more on this see e.g. [83, Chapter 7].

We will not prove the claim since it's a slight detour from our main goal of applying entropy to combinatorics problems. Most proofs can be found in standard texts on entropy, e.g. the survey by Galvin [74] (which contains even more properties, especially around conditional entropy). The only exceptions might be proofs for the "only if" portion of Dropping Conditioning (which we will never use); as well as the Data Processing Inequality (whose name is not entirely standard), but this follows from observing that $H[X|Y, Z] = H[X|Y]$ by definition and that $H[X|Y, Z] \geq H[X|Z]$ by Dropping Conditioning.

There is plenty of redundancy in the list of properties above. For example Subadditivity follows from the Chain Rule and Dropping Conditioning, but this property is used so frequently that it will be useful to list it as a separate property. For our purposes, the most useful property from this list will be the if and only if portion of the Maximality Principle. More precisely, we have the following consequence of it (our name for which is non-standard).

Corollary 11.2 (Fundamental Property of Entropy)**.** *If $\mathcal{X}$ is a set and $X \in \mathcal{X}$ is chosen uniformly at random, then*

$$|\mathcal{X}| = 2^{H[X]}.$$

This corollary shows that bounding the size of a set is equivalent to bounding the entropy of a corresponding random variable, and sometimes this latter perspective can be easier to work with, especially when it comes to proving upper bounds.

11.2 Basic Application 1: Binomial Coefficients

We begin with a typical usage for entropy by upper bounding (sums of) binomial coefficients. For this we define for $p \in [0,1]$ the binary entropy function

$$H(p) := -p\log_2(p) - (1-p)\log_2(1-p),$$

which is simply the entropy of a Bernoulli variable with probability of success p.

Proposition 11.3. *For all $k \le n/2$, we have*

$$\sum_{i=0}^{k} \binom{n}{i} \le 2^{H(k/n)\cdot n}.$$

Proof. Let $\mathcal{X}$ denote the set of binary strings of length n with at most k 1's. Note that

$$|\mathcal{X}| = \sum_{i=0}^{k} \binom{n}{i},$$

so by the Fundamental Property of Entropy, proving our desired bound $\log|\mathcal{X}| \le H(k/n)\cdot n$ is equivalent to showing $H[X] \le H(k/n)\cdot n$ where $X \in \mathcal{X}$ is chosen uniformly at random. For this we use the following.

Claim 11.4. *Each of the random variables X_i is a Bernoulli random variable with probability of success $p \le k/n \le 1/2$.*

Proof. By the symmetry of the problem, we see that each X_i is Bernoulli with the same probability of success p. By construction we deterministically have $\sum X_i \le k$ for all $X \in \mathcal{X}$, so it must be that $p \le k/n$, and this is at most $1/2$ by hypothesis on k. □

The claim above implies that for all i,

$$H[X_i] = H(p) \le H(k/n),$$

with this last step using the (easy to prove) fact that $H(p)$ is increasing for $p \leq 1/2$. This together with Subadditivity gives

$$H[X] \leq \sum H[X_i] = H(k/n) \cdot n,$$

proving the result. $\square$

We note that this bound is essentially tight in the sense that if k is linear in n, then $\binom{n}{k} = 2^{(1+o(1))H(k/n)\cdot n}$.

The framework in our proof above is typical for the entropy method: we (1) started with a uniform random object X and then (2) used entropy to upper bound the size of $\mathrm{supp}(X)$. Although this is the most common framework for using entropy, it is also possible to (1') start with a non-uniform random object X and then (2') use entropy to lower bound the size of $\mathrm{supp}(X)$. We study one such example in the next subsection.

11.3 Basic Application 2: Walks in Graphs

One of the most famous open problems in extremal combinatorics is Sidorenko's conjecture stated below, where we recall that $\hom(F, G)$ denote the number of homomorphisms from a graph F to a graph G.

Conjecture 11.5 (Sidorenko's Conjecture)**.** *If F is a bipartite graph, then every n-vertex graph G satisfies*

$$\frac{\hom(F, G)}{n^{|V(F)|}} \geq \left(\frac{2e(G)}{n^2}\right)^{e(F)}.$$

This bound is asymptotically best possible for $G = G_{n,p}$, so the conjecture morally says that every graph G contains as many (homomorphic) copies of each bipartite graph as one would expect in the random graph of the same density. Note that the bound does not hold for F non-bipartite (via taking $G = K_2$).

There are very few examples of graphs F for which Sidorenko's Conjecture is known. Here we outline a nice entropy proof which shows that Sidorenko's Conjecture holds for all F which are paths. In this setting, we observe that homomorphisms from paths of length k to G are equivalent to walks of length k in G, i.e. sequences of vertices $(x_1, \ldots, x_{k+1})$ such that $x_i \sim x_{i+1}$ for all $1 \leq i \leq k$. As such, the following result (commonly known as the Blakey-Roy Theorem (though they weren't quite the first to prove it)) is equivalent to Sidorenko's Conjecture for paths.

Theorem 11.6 (Blakey-Roy [24])**.** *If G is an n-vertex graph with $m \geq 1$ edges, then the number of walks of length k in G is at least*

$$2m(2m/n)^{k-2}.$$

Note that this bound is tight whenever G is regular.

Proof. Let $X = (X_1, \ldots, X_{k+1})$ be a random walk of length k in G chosen in the following *non-uniform* way: choose the pair (X_1, X_2) uniformly at random amongst all pairs such that $X_1 \sim X_2$, and given X_{i-1} for $i \geq 3$, we choose X_i uniformly at random amongst the neighbors of X_{i-1}. Observe that X is indeed always a walk of length k (with us implicitly using that (X_1, X_2) exists due to $m \geq 1$). We can express its entropy as

$$\begin{aligned} H[X] &= \sum_{i=1}^{k+1} H[X_i | X_{<i}] \\ &= H[X_1] + H[X_1|X_2] + \sum_{i=3}^{k+1} H[X_i|X_{i-1}] \\ &= H[X_1, X_2] + \sum_{i=3}^{k+1} H[X_i, X_{i-1}] - H[X_{i-1}], \end{aligned} \tag{11}$$

where here the first and last equality used the Chain Rule, and the second used the equality case of Dropping Conditioning (since X_i depends only on X_{i-1}).

We claim (crucially) that (X_{i-1}, X_i) is uniformly random amongst all pairs such that $X_{i-1} \sim X_i$. This is true for $i = 2$ by construction, so assume we have proven it true up to some value $i \geq 3$. In this case, for any $(y, z) \in V(G)^2$ such that $y \sim z$, we have (by conditioning on every possible value that $X_{i-2} \sim X_{i-1}$ can take on)

$$\begin{aligned} \Pr[(X_{i-1}, X_i) = (y, z)] &= \sum_{x \in N(y)} \Pr[(X_{i-2}, X_{i-1}) = (x, y)] \cdot \Pr[X_i = z | (X_{i-2}, X_{i-1}) = (x, y)] \\ &= \sum_{x \in N(y)} \frac{1}{2m} \cdot \frac{1}{\deg(y)} = \frac{1}{2m}, \end{aligned}$$

with this last equality using the hypothesis that (X_{i-2}, X_{i-1}) is distributed uniformly at random and that X_i is a uniform random neighbor of X_{i-1}. This establishes the claim.

With this claim, we have by the Maximization Principle that $H[X_i, X_{i-1}] = \log(2m)$ for all i, and also that $H[X_{i-1}] \leq \log(n)$ for all i. Using this with (11) gives

$$H[X] \geq (k-1)\log(2m) - (k-2)\log(n).$$

Let W_k denote the set of walks of length k in G. Since X is a random element from W_k, we have $H[X] \leq \log |W_k|$, which combined with the lower bound for $H[X]$ above gives the desired result. $\square$

To emphasize, proof above would theoretically have worked if we considered a uniform random element $\tilde{X} \in W_k$ instead of the non-uniform X described above, in the sense that the Maximization Principle gives

$$H[\tilde{X}] \geq H[X] \geq (k-1)\log(2m) - (k-2)\log(n).$$

However, it is not clear how one would prove the lower bound of $(k-1)\log(2m) - (k-2)\log(n)$ for $H[\tilde{X}]$ directly without going through X first. More generally, it can be useful when working with lower bounds from entropy to work with a distribution for X which is "natural" to the problem rather than one which is uniform. We also emphasize that to obtain tight examples, it is crucial that our choice of X is actually distributed uniformly whenever we are working with an extremal example (as otherwise the Maximization Principle shows that we can not hope to obtain a tight bound).

The exact same proof as above can easily be extended to prove Sidorenko's Conjecture for all trees at the cost of more complicated notation. A different and somewhat more involved entropy argument can be used to show Sidorenko's conjecture holds whenever F has a vertex which is adjacent to every vertex in the other partition set. This result was originally proven by Conlon, Fox, and Sudakov [41] using Dependent Random Choice **And I'm not sure who gave the entropy proof for this, though it can be found in notes of Yufei Zhao**.

11.4 Further Applications

While one can prove a number of results using only the entropy properties listed above (see for example Theorem 14.2), we can greatly increase the power of this method by adding in a few extra tools which we explore in the forthcoming two sections. Much more can be said regarding the entropy method than we do in this text, and we refer the reader to the survey by Galvin [74] for much more on this topic.

12 Shearer's Lemma

In this section we consider a strengthening of the Subadditivity property of entropy known as Shearer's lemma, which plays a central role in many applications of the entropy method (so much so that it is often listed as one of the basic properties of entropy). For this, given a vector $X = (X_1, \ldots, X_n)$ and a set $F = \{i_1, \ldots, i_k\} \subseteq [n]$, we define $X_F = (X_{i_1}, \ldots, X_{i_k})$.

Lemma 12.1 (Shearer's Lemma). *Let $X = (X_1, \ldots, X_n)$ and let $\mathcal{F} \subseteq 2^{[n]}$ be a set system such that every $i \in [n]$ is contained in at least d sets of $\mathcal{F}$, i.e. such that $\mathcal{F}$ has minimum degree at least d. Then*

$$H[X] \leq d^{-1} \sum_{F \in \mathcal{F}} H[X_F].$$

Note that if $\mathcal{F}$ consists of the singletons $\{i\}$, then Shearer's lemma with $d = 1$ exactly recovers subadditivity.

The statement of Lemma 12.1 is a little strange at first glance. Roughly speaking, the intuition is that because each i appears at least d times in $\mathcal{F}$, the sum on the righthand side in principle has as much information as d independent copies of X, and hence dividing by d should in principle be an upper bound for the amount of information of X. This intuition can easily be made precise if the X_i are mutually independent, as in this case we have

$$d \cdot H[X] = \sum_i d \cdot H[X_i] \leq \sum_i \deg_{\mathcal{F}}(i) \cdot H[X_i] = \sum_{F \in \mathcal{F}} H[X_F].$$

A similar line of reasoning gives the result in general.

Proof. Our proof will go through the same logic as in the independent case, except that we will replace both of our equalities with an application of the Chain Rule. To this end, given i, F with $i \in F \in \mathcal{F}$, we define $(i, F) = \{j \in F : j < i\}$. With this we have

$$\begin{aligned} d \cdot H[X] &= \sum_i d \cdot H[X_i | X_{<i}] \\ &\leq \sum_i \deg_{\mathcal{F}}(i) \cdot H[X_i | X_{<i}] = \sum_i \sum_F H[X_i | X_{<i}] \\ &\leq \sum_i \sum_F H[X_i | X_{(i,F)}] = \sum_F H[X_F], \end{aligned}$$

where this last inequality used Dropping Conditioning. $\square$

We begin with a basic application of Shearer's lemma to a problem from geometry. Given a vector $x \in \mathbb{Z}^n$ and $i \in [n]$, we define the projection $\pi_i(x) = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$, and given $S \subseteq \mathbb{Z}^n$ we define $\pi_i(S) = \{\pi_i(x) : x \in S\}$. A natural extremal question is: given the sizes of $\pi_i(S)$ for all i, how large can S itself be?

A natural candidate for a construction here is to take S to be a cube whose ith side length is $|\pi_i(S)|^{-(n-2)/(n-1)} \prod_j |\pi_j(S)|^{1/(n-1)}$. Indeed, it is easy to check that each $\pi_i(S)$ has the prescribed size, and with this we have $|S| = \prod |\pi_i(S)|^{1/(n-1)}$. It turns out this is best possible.

Theorem 12.2 (Discrete Loomis-Whitney Theorem). *If $S \subseteq \mathbb{Z}^n$, then*

$$|S| \leq \prod |\pi_i(S)|^{1/(n-1)}.$$

This result is essentially equivalent to the Loomis-Whitney Theorem [109], which says that if K is a measurable body in $\mathbb{R}^n$, then $\mathrm{vol}_n(K) \leq \prod \mathrm{vol}_{n-1}(\pi_i(K))^{1/(n-1)}$, where here π_i is the corresponding projection map and vol_n is the n-dimensional volume measure. Indeed, one can shows that it suffices to prove Loomis-Whitney when K is the union of axis-aligned unit cubes via an approximation argument, and in this case one can replace the center of each unit cube of K by its center, which exactly translates to the present problem.

Proof. As per usual, we start by picking X to be a uniformly random element of S. Observe that, in the notation of Shearer's lemma, $\pi_i(X) = X_{[n]\setminus\{i\}}$. Motivated by this, we define $\mathcal{F}$ to consist of all sets of the form $[n] \setminus \{i\}$. Since this set system has minimum degree $n-1$, Shearer's lemma gives

$$\log_2 |S| = H[X] \leq \frac{1}{n-1} \sum H[X_{[n]\setminus\{i\}}] = \frac{1}{n-1} \sum H[\pi_i(X)] \leq \frac{1}{n-1} \sum \log_2 |\pi_i(S)|,$$

with this last step using $\pi_i(X) \in \pi_i(S)$ (since $X \in S$) together with the Maximization Principle. Exponentiating both sides gives the result. $\square$

We next look at a more involved application of Shearer's Lemma involving homomorphisms between graphs. To this end, we let $\hom(G, H)$ denote the number of homomorphisms ϕ from G to H. For example, $\hom(G, K_r)$ is the number of proper r-colorings of G. We will also want to allow our target graph H to contain loops. For example, if H is a K_2 with a loop on one vertex, then one can check that $\hom(G, H)$ counts the number of independent sets of G.

The fundamental question we want to ask is: given a target graph H and some family of graphs $\mathcal{G}$, which $G \in \mathcal{G}$ maximizes $\hom(G, H)$? One natural family to consider is the set of n vertex d-regular graphs. While this problem is still open for general H, there is a very clean solution when we restrict to bipartite G.

Theorem 12.3 ([96, 147]). *If G is a bipartite n-vertex d-regular graph, then for any graph H (possibly with loops), we have*

$$\hom(G, H) \leq \hom(K_{d,d}, H)^{n/2d}.$$

This bound is sharp by considering G to be the disjoint union of $n/2d$ copies of $K_{d,d}$. This result was first proven in the case when H is a K_2 with a loop (i.e. when counting independent sets in G) by Kahn [96], with his argument later generalized considerably by Galvin and Tetali [76] to the present theorem.

Proof. It will be convenient to express our homomorphisms as vectors. To this end, given a homomorphism $\phi : V(G) \to V(H)$, we define the vector $X^\phi \in V(G)^{V(H)}$ by having $X^\phi_u = \phi(u)$ for each $u \in V(G)$. Let ϕ be a uniform random homomorphism from G to H, and let $X = X^\phi$.

Let $U \cup V$ denote a bipartition of G, noting that $|U| = |V| = n/2$ since G is regular. By the Maximization Principle and the chain rule, we have

$$\log(\hom(G, H)) = H[X] = H[X_U] + H[X_V|X_U].$$

Because G is d-regular, every $u \in U$ is contained in exactly d of the sets $\mathcal{F} := \{N(v) : u \in U\}$, so by Shearer's lemma we have

$$H[X_U] \leq d^{-1} \sum_{v \in V} H[X_{N(v)}].$$

By Subadditivity and dropping Conditioning we have

$$H[X_V|X_U] \leq \sum_{v \in V} H[X_v|X_U] \leq \sum_{v \in V} H[X_v|X_{N(v)}],$$

so in total we find

$$\log(\hom(G, H)) = H[X] \leq d^{-1} \sum_{v \in V} H[X_{N(v)}] + d \cdot H[X_v|X_{N(v)}], \tag{12}$$

and it now suffices to upper bound each term in the sum.

Fix some $v \in V$. For each vector x of length d indexed by $V(H)$, let $A(x)$ denote the set of vertices of $V(H)$ which are adjacent to every vertex of $x_1, \ldots, x_d$. Equivalently, $A(x)$ is the set of "available" values that X_v can take given that $X_{N(v)} = x$ (here we use that X corresponds to a homomorphism, i.e. X_v must be adjacent to every vertex of $X_{N(v)}$). We have $H[X_v|X_{N(v)} = x] \leq \log |A(x)|$ by the Maximization Principle, which together with the definition of (conditional) entropy gives

$$\begin{aligned} H[X_{N(v)}] + d \cdot H[X_v|X_{N(v)}] &= \sum_x \Pr[X_{N(v)} = x] \big(-\log(\Pr[X_{N(v)} = x]) + d \cdot H[X_v|X_{N(v)} = x]\big) \\ &\leq \sum_x \Pr[X_{N(v)} = x] \left(\log\left(\frac{|A(x)|^d}{\Pr[X_{N(v)} = x]}\right)\right) \\ &\leq \sum_x \log(|A(x)|^d) = \log\left(\prod_x |A(x)|^d\right), \end{aligned}$$

where the second inequality used Jensen's inequality together with the convexity of log.

Crucially, we observe that $\prod_x |A(x)|^d$ where x ranges all d-length vectors indexed of H is exactly equal to $\hom(K_{d,d}, H)$. Indeed, each homomorphism from $K_{d,d}$ can be uniquely identified by first choosing a vector x based on how the vertices on the bottom half of $K_{d,d}$ map to H, and after this each of the d remaining vertices of$K_{d,d}$ can map to any vertex in $A(x)$ while maintaining the homomorphism property. Combining this with the inequality above gives

$$H[X_{N(v)}] + d \cdot H[X_v|X_{N(v)}] \leq \log(\hom(K_{d,d}, H)),$$

and plugging this into (12) and using $|V| = n/2$ gives

$$\log(\hom(G, H)) \leq d^{-1} \sum_{v \in V} \log(\hom(K_{d,d}, H)) = \log(\hom(K_{d,d}, H))^{n/2d}.$$

□

The result above gives optimal bounds for $\hom(G, H)$ when G is bipartite, and an elegant idea of Zhao's allows one to immediately lift this result to all G for certain H. In particular, we get the following.

Theorem 12.4 (Zhao [147])**.** *Let $i(H)$ denote the number of independent sets of a graph H. If G is an n-vertex d-regular graph, then $i(G) \leq i(K_{d,d})^{n/2d}$.*

Proof Sketch. Consider the "bipartite double cover" $G \times K_2$, i.e. the graph with vertex set $V(G) \times \{1, 2\}$ where $(u, s) \sim (v, t)$ if and only if $u \sim_G v$ and $s \neq t$. A simple but clever injectivity argument shows $i(G \times K_2) \geq i(G)^2$. Because $G \times K_2$ is a d-regular $(2n)$-vertex bipartite graph, we have by Theorem 12.3 (applied when H is K_2 with a loop) that

$$i(G)^2 \leq i(G \times K_2) \leq i(K_{d,d})^{n/d},$$

giving the result. □

This same bipartite double cover trick can be used to extend Theorem 12.3 to all d-regular graphs for a few other choices of H [148], though there exist H for which such an extension is impossible (e.g. H being two isolated vertices with loops fails for $G = K_3$). Finally, we note a substantial extension of Theorem 12.3 due to Sah, Sawhney, Stoner, and Zhao [132], who showed that any graph G (not necessarily regular) which is triangle-free and has no isolated vertices satisfies for all graphs H that

$$\hom(G, H) \leq \prod_{uv \in E(G)} \hom(K_{d_u, d_v}, H)^{\frac{1}{d_u d_v}}.$$

There proof does not use entropy at all, but instead a very elaborate set of analytic inequalities.

13 Random Chain Rules

This section concerns a strengthening of the chain rule. Historically, this strengthening was first used to give an entropy proof of the following result.

Theorem 13.1 (Brégman's Theorem [28]). *If G is a bipartite graph with bipartition $U \cup V$ such that $|U| = |V| = n$, then the number of perfect matchings of G is at most*

$$\prod_{u \in U} (\deg(u)!)^{1/\deg(u)}.$$

Observe that this bound is tight by considering disjoint unions of complete bipartite graphs. This was originally proven by Brégman [28], and later Radhakrishnan [123] gave an elegant entropy-based proof that we present below.

Before stating the key lemma needed to prove Theorem 13.1, let's first try and prove this result naively from first principles and see where things go wrong.

As usual, we start with a uniform random perfect matching M of G. We then translate M into a vector X indexed by U by having $X_u \in V$ be the unique neighbor of u in M. We fix some arbitrary ordering $<$ of the vertices of U, and then apply the chain rule to obtain

$$H[X] = \sum_u H[X_u | X_{<u}].$$

At this point, the naive entropy argument calls for upper bounding $H[X_u|X_{<u}]$ by log of the number of possible values X_u can take given the values of $X_{<u}$. To this end, we define $A_{<u} \subseteq N(u)$ to be the set of neighbors of u which do not appear in $X_{<u}$ (i.e. this is the set of available neighbors of u for M given the information in $X_{<u}$). The Maximality Principle then gives

$$\sum_u H[X_u|X_{<u}] \leq \sum_u \log(|A_{<u}|).$$

Unfortunately, for any given u and ordering $<$, we can't say anything about $|A_{<u}|$ other than $|A_{<u}| \leq \deg(u)$. Applying this worst-case bound for all u gives a trivial upper bound of $\prod_u \deg(u)$ in the end.

While it is true that worst case we can have $|A_{<u}| = \deg(u)$ for any given u, intuitively we should "typically" have $|A_{<u}| \approx \frac{1}{2}\deg(u)$, since for a "random" M we would expect around half of u's neighbors to be matched in M to vertices appearing before u in $<$ and half to be matched to vertices after u. Again, this intuition may not hold for a given M and $<$, but this intuition can be made precise if we consider a random ordering $<$ instead of a fixed one. To this end, one can consider the following random variant of the chain rule.

Lemma 13.2 (Random Chain Rule). *Let X be a vector indexed by a set S and let $<$ be a random ordering of S. Then*

$$H[X] = \sum_{s \in S} \mathbb{E}_{<}[H[X_s|X_{<s}]].$$

Indeed, the proof of this follows from the fact that equality holds for any fixed $<$ (by the usual chain rule), and hence equality also holds when one takes expectations. With this we can quickly adapt our previous failed attempt to give Theorem 13.1.

Proof of Theorem 13.1. Let $<$ denote a uniform random ordering of U. Keeping all of the notation from the argument above, we have by the random chain rule that

$$H[X] = \sum_u \mathbb{E}_<[H[X_u|X_{<u}]] \leq \sum_u \mathbb{E}_<[\log(|A_{<u}|)]. \tag{13}$$

We claim that $|A_{<u}|$ is distributed uniformly at random amongst $[\deg(u)]$. Indeed, fix any perfect matching M (so now all the randomness lies in the much simpler random object $<$) and let v denote the neighbor of u in M. We then observe that $|A_{<u|}| = i$ if and only if $i-1$ vertices of $N(u) \setminus \{v\}$ have their neighbors in M appear after u under $<$. Since $<$ gives a uniform random ordering on $N(u)$ regardless of our choice of M, we conclude that $|A_{<u}|$ is indeed equally likely to be any value in $[\deg(u)]$. Again, we emphasize that this result holds regardless of the fixed value of M, and hence $|A_{<u}|$ continues to be uniformly distributed even if we do not condition on M.

With this claim, we can write (13) above as

$$H[X] \leq \sum_u \frac{1}{\deg(u)} \sum_{i=1}^{\deg(u)} \log(i) = \sum_u \frac{\log(\deg(u)!)}{\deg(u)}.$$

Exponentiating both sides gives the result. $\square$

Unpublished work of Kahn and Lovász generalizes Brégman's theorem to non-bipartite graphs by proving that every graph has at most $\prod_{u \in V(G)} d(u)!^{1/2d(u)}$ perfect matchings. An entropy proof of this result was given by Cutler and Radcliffe [45], with this proof being complicated by the fact that $|A_{<u}|$ is no longer uniformly distributed if u has edges inside $N(u)$. A short proof of Alon and Friedland [4] deduces this result of Kahn and Lovász directly from Brégman's theorem.

Our proof above used the random chain rule to count the number of 1-factors (i.e. perfect matchings) in a graph. We next use a somewhat more complex version of this argument to count 1-factorizations in K_n, which we recall are ordered partitions of $E(G)$ into perfect matchings. For example, K_4 has 6 different 1-factorizations, namely $(\{12, 34\}, \{13, 24\}, \{14, 23\})$ and all of its permutations.

I don't know if there's a relevant citation here.

Theorem 13.3. *The number of 1-factorizations of K_n when n is even is at most $((1 + o(1))n/e^2)^{\binom{n}{2}}$.*

Observe that this improves upon the trivial upper bound $(n-1)^{\binom{n}{2}}$ (which is just the number of ways to partition the edge set of K_n into $n-1$ edge-disjoint graphs).

Proof. For this proof, it will be slightly more convenient to work with a base e notion of entropy rather than the usual base 2 notion. To this end, if X is a random variable with $p_x = \Pr[X = x]$, then we define

$$H_e[X] = -\sum_x p_x \log_e(P_X).$$

Note that $H_e[X] = \log_e(2)H[X]$, and in particular, essentially all of the properties for $H[X]$ continue to hold for $H_e[X]$.

Let M denote a uniformly random 1-factorization of K_n. We will think of M as assigning to each edge uv of K_n a color in $[n-1]$ such that the edges in color i form a perfect matching (equivalenlty, M is a proper edge coloring of K_n with $n-1$ colors). Let X be the vector indexed by $E(K_n)$ where X_{uv} equals the color assigned to uv by M.

As before, we will consider a uniformly random ordering $<$ on $E(K_n)$, but for technical reasons we will want to form this ordering in a slightly more complex way. To this end, assign to each edge uv a random weight w_{uv} chosen independently and uniformly from $[0,1]$, then let $<$ be the ordering of $E(K_n)$ which has $uv < xy$ iff $w_{uv} < w_{xy}$. Again we emphasize that $<$ has the same distribution as if we just chose it to be uniformly at random, but it will be convenient for us to have these extra w_{uv} parameters to work with. An application of the random chain rule then gives

$$H_e[X] = \sum_{uv} \mathbb{E}_<[H_e[X_{uv}|X_{<uv}]] \le \sum_{uv} \mathbb{E}_<[\log_e(|A_{<uv}|)], \tag{14}$$

where here A_{uv} denotes the set of colors that are "available" for uv given $X_{<uv}$; i.e. A_{uv} consists of the colors i that do not lie on any edge xy which intersects uv and which has $xy < uv$ under the coloring M.

It remains to estimate $\mathbb{E}_<[\log_e(|A_{<uv}|)]$, and for this, it will suffice to condition on the 1-factorization M and prove an upper bound that is independent of M. From now on we fix M, noting that $A_{<uv}$ will always contain the color c which M assigns to uv. Observe that a color $i \in [n-1] \setminus \{c\}$ will be in $A_{<uv}$ if and only if uv appears in the ordering before the two edges incident to uv which are colored i by M (note that exactly two such edges exist since M is a 1-factorization). Conditional on the value w_{uv}, the probability that this happens for any given i is $(1-w_{uv})^2$. Thus in total, we have

$$\mathbb{E}_<[\log_e(|A_{<uv}|)|w_{uv}, M] = \log_e(1 + (n-2)(1-w_{uv})^2).$$

As w_{uv} was distributed uniformly at random in $[0,1]$, we find

$$\mathbb{E}_<[\log_e(|A_{<uv}|)|M] = \int_0^1 \log_e(1 + (n-2)(1-x)^2)dx.$$

Summing this over all edges uv together with (14) gives

$$H_e[X] \le \binom{n}{2} \int_0^1 \log_e(1 + (n-2)(1-x)^2)dx = \log_e(n) - 2 + o(1),$$

where this last equality follows from some fiddly integral analysis[1] (with the intuition being that the integrand is close to $\log_e(n(1-x)^2) = \log_e(n) + 2\log_e(1-x)$, and this ends up integrating to the desired value). Exponentiating both sides by e gives that the total number of 1-factorizations is at most $((1+o(1))n/e^2)^{\binom{n}{2}}$ as desired.

$\square$

[1]Slightly more precisely, one can argue that the integral evaluated from $1 - n^{-.1}$ to 1 is at most $O(n^{-.1}\log n) = o(1)$, and outside of this range the difference between the integrand $\log_e(1 + (n-2)(1-x)^2)$ and $\log_e(n(1-x)^2)$ is $o(1)$.

As an aside, if in (14) we used the Chain Rule instead of the Random Chain Rule, then we would pessimistically use $|A_{<uv}| \leq n-1$ for all uv, and this would give the trivial upper bound $(n-1)^{\binom{n}{2}}$.

14 The Union-Closed Sets Conjecture

Up to this point we've used entropy to bound the size of some set, and this is certainly the most common way to utilize entropy for combinatorial problems. However, there are more exotic ways that entropy can be used to solve problems. We illustrate one such example here involving a major breakthrough by Gilmer for the Union-Closed Sets Conjecture. For this, we say that a set system $\mathcal{F}$ is *union-closed* if for all $A, B \in \mathcal{F}$, we also have $A \cup B \in \mathcal{F}$. The following (frustratingly) simple extremal question related to union-closed families is typically attributed to Frankl from 1979 [64].

Conjecture 14.1 (Union-Closed Sets Conjecture)**.** *If $\mathcal{F} \neq \{\emptyset\}$ is union-closed, then there exists some x in at least half of the elements of $\mathcal{F}$, i.e. such that $\deg(x) \geq \frac{1}{2}|\mathcal{F}|$.*

We omit discussing the history of this conjecture prior to 2022 (most of which can be found in the survey by Bruhn and Schaudt [29]). The only thing we note is that after many years of work, the best bound known was that there exists an x with $\deg(x) = \Omega(\frac{|\mathcal{F}|}{\log_2(|\mathcal{F}|)})$ due to Knill [105]. This bound stood for nearly 30 years until the following breakthrough result.

Theorem 14.2 (Gilmer [77])**.** *There exists a constant $c > 0$ such that if $\mathcal{F} \neq \{\emptyset\}$ is a union closed set system, then there exists an element x such that $\deg(x) \geq c|\mathcal{F}|$.*

Gilmer originally proved this with $c = .01$, though his methods were quickly optimized to give $c = \frac{3-\sqrt{5}}{2} \approx .381$ by a variety of authors, and our particular approach for achieving this constant will be closest to that of Chase and Lovett [35].

The key insight of Gilmer's approach is to look at the contrapositive of the Union-Closed Sets Conjecture, i.e. by showing that if a set system has $\deg(x) < c|\mathcal{F}|$ for all x, then $\mathcal{F}$ can not be union-closed. In particular, we will reach this union-closed conclusion through the following entropy result.

Theorem 14.3. *There exists a constant $c > 0$ such that the following holds. If $\mathcal{F}$ is a set system with $|\mathcal{F}| > 1$ and the property that $\deg(x) < c|\mathcal{F}|$ for all x, and if $A, B \in \mathcal{F}$ are chosen independently and uniformly at random, then*

$$H[A \cup B] > H[A].$$

Let us assume this result for the moment and show how it implies Gilmer's theorem.

Proof of Theorem 14.2. Let $\mathcal{F} \neq \{\emptyset\}$ be union-closed. The result is trivial if $\mathcal{F}$ has only 1 element, so assume this is not the case.

Let $A, B \in \mathcal{F}$ be chosen independently and uniformly at random. Because $\mathcal{F}$ is union-closed, the random variable $A \cup B$ always lies in $\mathcal{F}$. As such, the Maximality Principle implies.

$$H[A \cup B] \leq \log|\mathcal{F}| = H[A],$$

with this last step using that A is chosen uniformly at random. This contradicts the conclusion of Theorem 14.3, so it must be the case that $\deg(x) \geq c|\mathcal{F}|$ for some element x. $\square$

The proof above illustrates that the Union-Closed Sets Conjecture would follow if Theorem 14.3 held with $c = .5$. However, it turns out that $c = \frac{3-\sqrt{5}}{2}$ is the best constant one can prove here.

We now move onto our proof of Theorem 14.3. To start, we will attempt to build some intuition by working out how one might go about trying to solve this problem, though the eager reader is welcome to jump ahead to Lemma 14.4 for the formal details.

One way you might think of proving Theorem 14.3 is through induction on the ground set of $\mathcal{F}$. To this end, we can think of A, B as characteristic vectors in $\{0,1\}^n$. In this case, the chain rule implies that the conclusion of Theorem 14.3 is equivalent to having

$$\sum_{i \le n} H[(A \cup B)_i | (A \cup B)_{<i}] > \sum_{i \le n} H[A_i | A_{<i}].$$

Given this formulation, one might hope to prove this inequality term by term, and indeed this is what we will do.

In particular, when $i = 1$ we have that A_1, B_1 are just Bernoulli random variables with some common probability of failure[1] p. In this case, $(A \cup B)_1$ is a Bernoulli random variable with probability of failure p^2. As such, we need to show that if p is large (i.e. if the element 1 fails to be in a large proportion of edges of $\mathcal{F}$), then

$$H(p^2) > H(p),$$

where here $H(x) := -x\log_2(x) - (1-x)\log_2(1-x)$ is the entropy function for a Bernoulli random variable with probability x. Intuitively, this inequality will hold if and only if p^2 is closer to $1/2$ than p is to $1/2$, so the cutoff point should be when these two distances from $1/2$ are the same. That is, we need to solve for $.5 - p^2 = p - .5$, or equivalently to compute the roots of $p^2 + p - 1 = 0$. One can quickly check that the positive root of this is $p = \phi := \frac{\sqrt{5}-1}{2}$.

In conclusion, this first step of our induction (heuristically) seems to hold provided $p > \phi$ and fails otherwise. For our full proof, we will in fact need the following generalization of the claim that $H(p^2) > H(p)$ when $p > \phi$.

Lemma 14.4. *For all $p, q \in [0,1]$, we have*

$$H(pq) \ge \frac{1}{2\phi}(pH(q) + qH(p)),$$

where again $H(x) := -x\log_2(x) - (1-x)\log_2(1-x)$ and $\phi := \frac{\sqrt{5}-1}{2}$.

The proof of this purely analytical lemma is somewhat annoying to prove, so we'll omit it from the text here; see [27, 35] for the details. With this lemma and our approach outlined above, we'll quickly be able to derive the following.

Lemma 14.5 ([35]). *If $A, B \in \{0,1\}^n$ are independent random variables with $\Pr[A_i = 0], \Pr[B_i = 0] \ge p$ for all i, then*

$$H[A \cup B] \ge \frac{p}{2\phi}(H[A] + H[B]).$$

[1] We let p denote the probability of failure here (rather than the convention of having p denote the probability of success) in order to make some of the later algebra cleaner.

Proof. By the Chain Rule and Data Processing Inequality, we have

$$H[A \cup B] = \sum_{i \le n} H[(A \cup B)_i | (A \cup B)_{<i}] \ge \sum_{i \le n} H[(A \cup B)_i | A_{<i}, B_{<i}]. \tag{15}$$

Fix some $1 \le i \le n$. For each $x \in \{0,1\}^i$, define $p(x) = \Pr[A_i = 0 | A_{<i} = x]$ and $q(x) = \Pr[B_i = 0 | B_{<i} = x]$. By Lemma 14.4 and the independence of A, B, we have for all x, y that

$$H[(A \cup B)_i | A_{<i} = x, B_{<i} = y] = H(p(x)q(y)) \ge \frac{1}{2\phi}\left(p(x)H(q(y)) + q(y)H(p(x))\right).$$

Multiplying this inequality by $\Pr[A_{<i} = x,\ B_{<i} = y] = \Pr[A_{<i} = x] \cdot \Pr[B_{<i} = y]$ and summing over all x, y gives

$$\begin{aligned} H[(A \cup B)_i | A_{<i}, B_{<i}] &\ge \frac{1}{2\phi}\Big(\mathbb{E}_{A_{<i}, B_{<i}}[p(x)H(q(y))] + \mathbb{E}_{A_{<i}, B_{<i}}[q(y)H(p(x))]\Big) \\ &= \frac{1}{2\phi}\left(\mathbb{E}_{A_{<i}}[p(x)] \cdot \mathbb{E}_{B_{<i}}[H(q(y))] + \mathbb{E}_{B_{<i}}[q(y)] \cdot \mathbb{E}_{A_{<i}}[H(p(x))]\right) \\ &= \frac{1}{2\phi}(\Pr[A_i = 0] \cdot H[B_i | B_{<i}] + \Pr[B_i = 0] \cdot H[A_i | A_{<i}]) \\ &\ge \frac{p}{2\phi}(H[B_i | B_{<i}] + H[A_i | A_{<i}]), \end{aligned}$$

where the second equality used the definition of conditional entropy and the definition of $q(y)$, and the last inequality used the hypothesis $\Pr[A_i = 0], \Pr[B_i = 0] \ge p$. Combining this with (15) gives

$$H[A \cup B] \ge \frac{p}{2\phi} \sum_{i \le n} H[A_i | A_{<i}] + H[B_i | B_{<i}] = \frac{p}{2\phi}(H[A] + H[B]),$$

with this last equality using the Chain Rule. □

We can now complete the proof of Theorem 14.3 with $c = 1 - \phi = \frac{3-\sqrt{5}}{2} \approx .381$.

Proof of Theorem 14.3. Let $c = 1 - \phi$ and identify $A, B, A \cup B$ by their characteristic vectors in $\{0,1\}^n$. Note that our hypothesis $|\mathcal{F}| > 1$ implies $H(A), H(B) > 0$. Because $\deg(x) < c|\mathcal{F}|$ for all x, we have $\Pr[A_i = 0], \Pr[B_i = 0] > 1 - c = \phi$ for all i. Hence the previous lemma (together with $H(A), H(B) > 0$) implies

$$H[A \cup B] > \frac{1}{2}(H[A] + H[B]) = H[A],$$

proving the result. □

Before moving on, let us take a few moments to comment on some strengthenings of Theorem 14.2. As noted earlier, it turns out that $c = \frac{3-\sqrt{5}}{2}$ is the best c one can take so that Theorem 14.3 remains true. Moreover, this value of c (rather than $c = 1/2$) turns out to be the optimal value for an analogous "approximate" version of the union closed conjecture, i.e. a variant where we only require that most of the unions $A \cup B$ lie in $\mathcal{F}$; see [35] for more on this.

Given the above, some sort of genuinely new approach is needed to get beyond this $c = \frac{3-\sqrt{5}}{2}$ barrier for the original Union-Closed Sets Conjecture, and such an approach was found by Sawin [133]. In essence, his idea is to look at three random sets $A, B, C \in \mathcal{F}$ such that B is independent of A and C but (crucially) A, C are correlated in some way. Again, $\mathcal{F}$ being union closed implies that we must have $H[A \cup B], H[A \cup C] \leq H[A]$, and hence that

$$\alpha H[A \cup B] + (1-\alpha) H[A \cup C] \leq H[A]$$

for all α. Sawin then argues that this inequality fails to hold for some α if every element is in at most a $\frac{3-\sqrt{5}}{2} + \epsilon$ proportion of the elements of $\mathcal{F}$, giving the desired improvement. Cambie [33] obtained the best bounds one can get via this approach of Sawin, showing that one can take $c \approx .38235$ (cf the previous value of $c = \frac{3-\sqrt{5}}{2} \approx .38197$).

Part V

Hypergraph Containers

This part is heavily based off of lecture notes by Balogh [12]. Throughout this section we let $\mathcal{I}(H)$ denote the set of independent sets of a hypergraph H and $\mathcal{I}_m(H)$ the set of independent sets of size m. We adopt the notation $\binom{n}{\leq k}$ to denote the number of subsets of $[n]$ of size at most k. Many of the bounds in this part will be rough approximations to the truth in order to emphasize the intuition of the results and techniques rather than the nitty gritty detail that is actually required.

15 Introduction

Many problems in extremal combinatorics can be stated in terms of independent sets of hypergraphs. For example, one can define $\mathcal{H}_n^{AP}$ to be the 3-graph on $[n]$ where every triple $S \subseteq [n]$ is a hyperedge if and only if S is a 3-term arithmetic progression. Thus Roth's theorem is equivalent to saying $\alpha(\mathcal{H}_n^{AP}) = o(n)$. Similarly one can define $\mathcal{H}_n^{\Delta}$ to be the 3-graph whose vertex set is $E(K_n)$ and whose hyperedges are triples of edges in K_n which form a triangle. Independent sets of $\mathcal{H}_n^{\Delta}$ are triangle-free subgraphs of K_n, so Mantel's theorem says $\alpha(H_n^{\Delta}) = \lfloor n^2/4 \rfloor$.

This part is dedicated to a powerful method of upper bounding the size of $\mathcal{I}(H)$. Observe that for any hypergraph H we have

$$2^{\alpha(H)} \le |\mathcal{I}(H)| \le \binom{n}{\alpha(H)} 2^{\alpha(H)} \le (2n)^{\alpha(H)}.$$

In particular, the upper bound follows because every independent set is a subset of a set of size $\alpha(H)$. More generally, we say that a collection $\mathcal{C}$ of subsets $C \subseteq V(H)$ is a set of *containers* for H if every independent set $I \in \mathcal{I}(H)$ is a subset of some $C \in \mathcal{C}$. If such a set of containers exists, then

$$|\mathcal{I}(H)| \le \sum_{C \in \mathcal{C}} 2^{|C|} \le |\mathcal{C}| 2^{\max_{C \in \mathcal{C}} |C|}. \tag{16}$$

Thus we will get an effective upper bound on $|\mathcal{I}(H)|$ whenever we can find a small collection of containers, each of which are relatively small. Sometimes we will be interested in finding the number of independent sets of H of size m. The same reasoning as above gives

$$\binom{\alpha(H)}{m} \le |\mathcal{I}_m(H)| \le |\mathcal{C}| \binom{\max_{C \in \mathcal{C}} |C|}{m}. \tag{17}$$

The method of hypergraph containers gives a systematic way of obtaining such a collection of containers whenever H satisfies some fairly mild conditions. The main condition we need is that the codegrees of H to be relatively small, and in practice this often corresponds to having some notion of supersaturation.

16 Graph Containers

While the general method of containers involves bounding independent sets of *hypergraphs*, one can get pretty far by only considering independent sets of *graphs*. To this end we prove the following graph container lemma, which will be the main workhorse for the rest of this section. Recall that a collection $\mathcal{C}$ of subsets $C \subseteq V(G)$ is a set of *containers* for G if every independent set $I \in \mathcal{I}(G)$ is a subset of some $C \in \mathcal{C}$.

Lemma 16.1. *Let G be an n-vertex graph and $t > 0$ a positive number. There exists a collection $\mathcal{C}$ of containers such that*

(a) $|\mathcal{C}| \le \binom{n}{\le n/t}$.

(b) $\Delta(G[C]) < t - 1$ *for all* $C \in \mathcal{C}$.

In other words, there exists a small set of containers $\mathcal{C}$ such that each $C \in \mathcal{C}$ is "small" in the sense that it induces a graph with small maximum degree.

Proof. Our proof will be algorithmic: we construct a (deterministic) algorithm which takes as input a set $I \subseteq V(G)$ and which outputs a pair $(S(I), A(I))$ such that $S(I) \subseteq I \subseteq S(I) \cup A(I)$, and we will ultimately use $\{S(I) \cup A(I) : I \in \mathcal{I}(G)\}$ as our set of containers. We now describe the algorithm.

Fix an arbitrary ordering of $V(G)$. As input we take in an independent set $I \subseteq V(G)$. We initially set $S = \emptyset$ and $A = V(G)$ (the former corresponds to a set of "selected" vertices which are in I, and the latter to the set of "available" vertices which could possibly be in I given the current stage of the algorithm). The algorithm proceeds as follows:

Step 1 If $\Delta(G[A]) < t - 1$, output $(S(I), A(I))$. Otherwise proceed to Step 2.

Step 2 Let v be the vertex of maximum degree in $G[A]$, with ties being broken according to the fixed ordering of $V(G)$. If $v \notin I$, then set $A = A - v$ and repeat Step 1. Otherwise proceed to Step 2.

Step 3 Set $A = A - v - N_{G[A]}(v)$, $S = S \cup \{v\}$. Proceed to Step 1.

Let's reiterate what's going on here. It's not difficult to show inductively that we always have $I \subseteq S \cup A$, so $S \cup A$ serves as a container set for I, and we would like to trim this set down as much as possible. We do this by selecting a vertex $v \in A \cap I$ and adding it to S. If v has large degree in $G[A]$, then v being in the independent set I means that its many neighbors are not, so we get to remove all of these vertices from A while maintaining $I \subseteq S \cup A$. In particular, since we keep going so long as $G[A]$ has large maximum degree, we know at each step of this process that we're removing many vertices from A.

Define

$$\mathcal{C} = \{S(I) \cup A(I) : I \in \mathcal{I}(G)\},$$

which is a set of containers since $I \subseteq S(I) \cup A(I)$ at every step of the algorithm. Since we terminate the algorithm precisely when $\Delta(G[A(I)]) = \Delta(G[S(I) \cup A(I)]) < t-1$ (the equality holds since $S(I)$ has no neighbors in $S(I) \cup A(I)$), (b) holds. It thus remains to verify (a). To do this, we note the following which is easy to verify.

Claim 16.2. *Let I_1, I_2 be two independent sets and let $(S_1, A_1), (S_2, A_2)$ be their outputs from the algorithm. If $S_1 = S_2$, then $A_1 = A_2$.*

This claim implies that given $S(I)$, the container $S(I) \cup A(I)$ is uniquely determined[1]. In particular, if we always have $|S(I)| \leq n/t$, then the number of containers will be at most $\binom{n}{\leq n/t}$. And indeed, each round of the algorithm has $\Delta(G[A]) \geq t-1$, so every time a vertex is added to S at least $1 + (t-1) = t$ vertices are removed from A. In particular, at most n/t vertices can be added to S, giving the result. □

Actually, a closer inspection of the proof gives the following.

Lemma 16.3. *Let G be a graph on n vertices and $t \in \mathbb{R}$. There is a collection $\mathcal{C}$ of containers and functions*

$$f : \mathcal{I}(G) \to \binom{V(G)}{\leq n/t}, \qquad g : \binom{V(G)}{\leq n/t} \to \mathcal{C}$$

such that the following hold.

(a) The function g is a surjection. In particular, $|\mathcal{C}| \leq \binom{n}{\leq n/t}$.

(b) We have $\Delta(G[C]) < t-1$ for all $C \in \mathcal{C}$.

(c) For every $I \in \mathcal{I}(G)$ we have

$$f(I) \subseteq I \subseteq g(f(I)).$$

Proof. Consider the exact same algorithm as before. Define $f(I) = S(I)$ and $g(S) = C(S)$ (if $S \neq S(I)$ for any I, then assign g arbitrarily). It's not hard to check that this works. □

The extra source of power of this lemma is that for each $I \in \mathcal{I}$ we are given some set $S = f(I)$ contained in I. In many examples this extra information is needed to get tight upper bounds when counting independent sets, though for pedagogical purposes we will often work with the simpler Lemma 16.1 to get close to tight results.

In the coming subsections we'll show how to use Lemma 16.1 to solve several combinatorial problems. All of the proofs will be very similar to each other, though they'll become increasingly sophisticated as we go along.

Before going on, let us briefly note that there are many variants of Lemma 16.1 that one can prove using a similar approach. These variants of Lemma 16.1 are both a blessing and a curse since they give many options for how to solve a given problem (and it isn't always clear which is best).

[1] Because of this, S is often called a "certificate" or "fingerprint" of I.

16.1 Regular Graphs

Our first application of Lemma 16.1 will be to count the number of independent sets in a d-regular graph. As a point of reference, it is not difficult to show that if G consists of $n/2d$ disjoint copies of $K_{d,d}$, then

$$|\mathcal{I}(G)| = (2^{d+1} - 1)^{n/2d} = 2^{n/2+n/2d+o(n)}.$$

Thus for d-regular graphs, we can't possibly hope to prove an upper bound on $|\mathcal{I}(G)|$ stronger than roughly $2^{n/2}$ when d is large. We can prove that this is close to best possible using containers.

Theorem 16.4. *Let G be a d-regular n-vertex graph with $\log n \ll d \ll n/2$. Then*

$$|\mathcal{I}(G)| \leq 2^{n/2+o(n)}.$$

In fact, it turns out that $|\mathcal{I}(G)| \leq (2^{d+1} - 1)^{n/2d}$ for all d-regular n-vertex graphs. This was proven for bipartite graphs by Kahn [96] using entropy, and the problem was solved in full by Zhao [147]. As far as I'm aware, the proof of Theorem 16.4 presented here is due to Balogh.

As a first step to proving Theorem 16.4, we will apply Lemma 16.1 to our graph G to get a collection of containers $\mathcal{C}$. We would like to conclude the result by the observation from (16):

$$|\mathcal{I}(G)| \leq \sum_{C \in \mathcal{C}} 2^{|C|} \leq |\mathcal{C}| 2^{\max_{C \in \mathcal{C}} |C|},$$

but there's an issue with this. Namely, Lemma 16.1 only tells us that each $C \in \mathcal{C}$ induces a graph in G with small maximum degree. For a general graph this tells us nothing about $|C|$, but fortunately in d-regular graphs, $G[C]$ having small maximum degree is only possible if C is small. The following states a precise version of the contrapositive of the previous sentence.

Lemma 16.5. *For any $\epsilon > 0$, if G is a d-regular graph and $C \subseteq V(G)$ with $|C| = n/2 + \epsilon n$, then $\Delta(G[C]) \geq 2\epsilon d$.*

This lemma is a form of supersaturation: a d-regular graph can have a subset of size $n/2$ with $G[C]$ empty (e.g. if G is bipartite), but if C is just a bit larger than this, then it must have relatively high maximum degree. As we will see, supersaturation results are almost always a necessary ingredient for applying the method of containers.

Proof. Because the maximum degree is always at least the average degree, we have

$$\Delta(G[C]) \geq 2e(G[C])/|C| \geq 2e(G[C])/n$$

, so it will suffice to show that $e(G[C])$ is large. To do this, we let $\overline{C} = V(G) \setminus C$ and note that

$$d|C| = \sum_{v \in C} d(v) = 2e(G[C]) + e(C, \overline{C}) \leq 2e(G[C]) + d|\overline{C}|.$$

Because $|C| = n/2 + \epsilon n$ and $|\overline{C}| = n/2 - \epsilon n$, in total this implies

$$2e(G[C]) \geq 2\epsilon dn.$$

Combining this with the observation at the start gives the result. $\square$

Corollary 16.6. *For all t, if G is an n-vertex d-regular graph, then there exists a set of containers $\mathcal{C}$ with $|\mathcal{C}| \le \binom{n}{\le n/t}$ and $|C| \le \frac{1}{2}n + \frac{t}{d}n$ for all $C \in \mathcal{C}$.*

Proof. Let $\mathcal{C}$ be a set of containers as guaranteed by Lemma 16.1. Because $\Delta(G[C]) < t-1 \le t$, Lemma 16.5 implies that $|C| \le \frac{1}{2}n + \frac{t}{d}n$. □

With this we can prove Theorem 16.4.

Proof of Theorem 16.4. At this point all we need to do is use (16) after applying Corollary 16.6 with a carefully chosen value of t. Note that

$$|\mathcal{C}| \approx \binom{n}{n/t} \approx 2^{n\log(t)/t},$$

and we already know $2^{\max|C|} \approx 2^{\frac{1}{2}n+\frac{t}{d}n}$. Thus to minimize $|\mathcal{C}| \cdot 2^{\max|C|}$ we should choose t so that $\frac{t}{d} \approx \log(t)/t$, and in particular $t = \sqrt{d\log n}$ is a reasonable choice. One can verify with a more formal argument that this does indeed give the desired result after applying (16). □

We note that the statement of Corollary 16.6 and the optimization of t in the proof of Theorem 16.4 is in some sense independent[1] of the problem of determining $|\mathcal{I}(G)|$ for G a d-regular graph. That is, these results are effective for other problems which involve counting independent sets of d-regular graphs.

For example, recall that a q-coloring of a graph G is a map $\chi : G \to [q]$ such that $\chi(u) \neq \chi(v)$ whenever $uv \in E(G)$. Equivalently, a q-coloring is a partition of $V(G)$ into independent sets $I_1, \ldots, I_q$, With this latter formulation, we can use containers to get an effective bound on the number of q-colorings of G, which we'll denote by $X_q(G)$.

Again, let's consider a test case to figure out how strong of a bound we could possibly hope to prove. Let G be $n/2d$ disjoint copies of $K_{d,d}$. We know that G has close to as many independent sets as it could possible have, so it seems plausible that it would have many q-colorings as well. In particular, one can prove that $X_q(G) \approx (q/2)^n$, and once again we can prove that this is essentially best possible.

Theorem 16.7 ([75])**.** *Let G be an n-vertex d-regular graph and q an integer such that $q^2 \log n \ll d$. Then*

$$X_q(G) \le (q/2 + o(1))^n.$$

We note that a stronger result was proven by Galvin [75] with a somewhat more involved proof.

Proof. By the same reasoning as in Theorem 16.4, there exists a set of containers $\mathcal{C}$ for G such that $|\mathcal{C}| \approx 2^{\sqrt{\frac{\log n}{d}}n}$ and $|C| \approx \frac{1}{2}n$ for each $C \in \mathcal{C}$. Consider all vectors of the form $(C_1, \ldots, C_q)$ with $C_i \in \mathcal{C}$, noting that the number of such vectors is at most $|\mathcal{C}|^q = 2^{o(n)}$.

Observe that every q-coloring can be identified by a vector $(I_1, \ldots, I_q)$ where each I_j is an independent set and $\bigcup I_j = V(G)$. Each of these vectors is "contained" in some "container

[1] Ha.

vector" $(C_1, \ldots, C_q)$ with $C_j \in \mathcal{C}$ in the sense that $I_j \subseteq C_j$ for all j. Thus it's enough to count how many q-colorings each container vector contains.

A naive upper bound for the number of q-colorings contained in $(C_1, \ldots, C_q)$ is roughly $2^{qn/2}$ since this is the number of ways to choose an independent set from each C_i. This bound is too weak, so we have to utilize the extra information that the I_j partition $V(G)$.

To this end, assume $V(G) = \{v_1, \ldots, v_n\}$. Given $(C_1, \ldots, C_q)$, let a_i be the number of containers C_j with $v_i \in C_j$. It's not difficult to see that the number of q-colorings contained in this vector is then at most $\prod a_i$, and by the AMGM inequality this is at most $(\sum a_i/n)^n = (\sum |C_j|/n)^n$. Each of the q containers has size at most roughly $n/2 + \sqrt{\frac{\log n}{d}}n$, so this gives the desired result. $\square$

16.2 A Randomized Sperner's Theorem

Throughout this subsection we fix an integer n and define $N := 2^n$ and $m = \binom{n}{\lfloor n/2 \rfloor}$. An *antichain* of $[n]$ is a subset $S \subseteq 2^{[n]}$ such that $A \not\subseteq B$ for any distinct $A, B \in S$. For example, $\binom{[n]}{k}$ is an antichain for all k. A famous result of Sperner's says that an antichain of $[n]$ has size at most m.

Our first goal is to count the number of antichains of $[n]$. To do this, we form a graph where independent sets correspond to antichains. Let G_N denote the graph whose vertex set is $2^{[n]}$ and where A, B are adjacent to each other if either $A \subseteq B$ or $B \subseteq A$. Analogous to Lemma 16.5, we need a supersaturation lemma for G_N which says that any collection of vertices that is much larger than m induces many edges. In particular, the following suffices.

Lemma 16.8 ([102, 15]). *If $C \subseteq 2^{[n]}$ has $|C| > (1+\epsilon)m$ with $0 < \epsilon \leq 1/3$, then $e(G_N[C]) \geq \epsilon mn/2$.*

We won't prove this, but we will briefly comment on some intuition for the result. Intuitively, if you want to build a set of size $(1+\epsilon)m$ which induces few edges, then a good place to start is with the middle layer $\binom{[n]}{\lfloor n/2 \rfloor}$ since this is a maximum independent set. From there one could greedily choose ϵm sets which have as few neighbors as possible in this middle layer, and in particular choosing them allfrom $\binom{[n]}{\lfloor n/2 \rfloor + 1}$ gives a total of $(\lfloor n/2 \rfloor + 1) \cdot \epsilon m \geq \epsilon mn/2$ edges. Kleitman [102] proved that this is indeed the best construction, and the exact numerical computation[1] was done by Balogh, Mycroft, and Treglown [15].

With our supersaturation lemma in hand, we can easily prove the following result of Kleitman.

Theorem 16.9 ([103]). *The number of antichains of $[n]$ is $2^{m+o(m)}$.*

Proof. We obtain a set of containers $\mathcal{C}$ for G_N by applying lemma 16.1 to G_N with a parameter t to be determined later. Let ϵ be such that $(1+\epsilon)m = \max_{C \in \mathcal{C}} |C|$ and let C be a container

[1]The numbers in Lemma 16.8 are slightly different from those that appear in [15], but it's not difficult to refine their proof to give this result.

achieving this bound. By Lemma 16.8 we have[1]

$$\Delta(G_N[C]) \geq 2e(G_N[C])/|C| \geq \epsilon mn/(1+\epsilon)m \approx \epsilon n.$$

By assumption this quantity is at most t, or equivalently we roughly have $\max_{|C| \in \mathcal{C}} |C| \leq (1+t/n)m$. By (16) we have an upper bound of roughly

$$\binom{N}{N/t} 2^{(1+t/n)m} \approx 2^{\frac{N}{t}\log(t)+(1+t/n)m}.$$

This quantity is optimized when $N\log(t)/t \approx tm/n$. We have $m \approx N/\sqrt{n}$, so in total we want $t \approx n^{3/4}/\sqrt{\log n}$, and one can verify that this choice of t gives the desired bound. $\square$

We next prove a random version of Sperner's theorem. The setup is as follows. Choose a subset $R_p \subseteq 2^{[n]}$ by including each set in R_p independently and with probability p. How large is the size of a largest antichain in R_p (in expectation)? In terms of our graph G_N, this is equivalent to computing $\mathbb{E}[\alpha(G_N[R_p])]$.

Let's consider some simple cases first. If $p = 1$, then $G_N[R_p] = G_N$ and we know its independence number is m. Somewhat more generally, we always have the lower bound $\mathbb{E}[\alpha(G_N[R_p])] \geq pm$ since this is the expected size of the set $\binom{[n]}{\lfloor n/2 \rfloor} \cap R_p$. However (as will often be the case), the behavior of $\mathbb{E}[\alpha(G_N[R_p])]$ changes considerably when p is very small.

For example, if $p^2 3^n \ll p 2^n$, then asymptotically we have $\mathbb{E}[\alpha(G_N[R_p])] \sim |R_p|$ by a simple deletion argument. Even above the deletion threshold it is possible to improve on the trivial lower bound. Indeed, construct an independent set I by keeping each vertex in $\binom{[n]}{\lfloor n/2 \rfloor} \cap R_p$ together with all the vertices in $\binom{[n]}{\lfloor n/2 \rfloor - 1} \cap R_p$ which are not contained in any of the vertices of $\binom{[n]}{\lfloor n/2 \rfloor} \cap R_p$. The expected number of vertices we get from this first part is $p\binom{n}{\lfloor n/2 \rfloor}$, and from the second is[2] $p(1-p)^{n+1-\lfloor n/2 \rfloor}\binom{n}{\lfloor n/2 \rfloor - 1}$. In particular, if $p = c/n$ for a fixed constant c, then this asymptotically gives $(1+\epsilon)pm$ for some $\epsilon > 0$.

It turns out that for larger p we do have $\mathbb{E}[\alpha(G_N[R_p])] \sim pm$. More precisely we have the following due to Balogh, Mycroft, and Treglown [15].

Theorem 16.10 ([15]). *For any $\epsilon > 0$, there exists a constant c so that if $p > c/n$, then a.a.s. $\alpha(G_N[R_p]) \leq (1+\epsilon)pm$.*

Roughly speaking the approach we would like to use is as follows. Observe that $\alpha(G_N[R_p]) \geq k$ if and only if $G_N[R_p]$ contains an independent set of size k. The expected number of such sets in $G_N[R_p]$ is exactly $p^k \mathcal{I}_k(G_N)$, and if this quantity is small then we can conclude the result by Markov's inequality. Thus to solve this problem (and in general to solve extremal problems in random sets), we need to get effective upper bounds on $\mathcal{I}_k(G_N)$.

Unfortunately a naive application of Lemmas 16.1 and 16.8 together with (17) turns out to be too weak. The bottleneck here is the supersaturation result from Lemma 16.8. While the

[1]Implicitly this assumes $\epsilon \leq 1/3$. One can get around this by taking $C' \subseteq C$ a set of size exactly $4m/3$, but ultimately this computation is just to obtain intuition for what value t should be.

[2]To keep a vertex it has to be in R_p and all its neighbors in $\binom{[n]}{\lfloor n/2 \rfloor - 1}$ have to be out.

stated bound is essentially tight for $|C| = (1+\epsilon)m$ with $0 < \epsilon \le 1/3$ and n sufficiently large, the bound is not tight when e.g. $|C| = (2+\epsilon)m$. Intuitively in this case the extremal example should come from taking two middle layers together with ϵm of the layer right above these two. In particular, each of the ϵm vertices will have degree about n^2, so we expect around ϵmn^2 edges in $G_N[C]$. And indeed, this is the case.

Lemma 16.11 ([102, 15]). *If $C \subseteq 2^{[n]}$ has $|C| > (2+\epsilon)m$ with $0 < \epsilon \le 1/3$, then $e(G_N[C]) \ge \epsilon mn^2/9$.*

With this we can prove the main result.

Proof of Theorem 16.10. The key idea is to apply the container lemma twice using the two different levels of supersaturation from Lemmas 16.8 and 16.11. In particular, let $\mathcal{C}_1$ be a set of containers coming from Lemma 16.1 using G_N and some t_1, and for each $C_1 \in \mathcal{C}_1$, let $\mathcal{C}_2(C_1)$ be a set of containers coming from Lemma 16.3 using $G_N[C_1]$ and some t_2.

We first want to choose t_1 so that each $C_1 \in \mathcal{C}_1$ has size roughly $\alpha(G_N) = m$. Observe that if $|C_1| > 3m$, then by Lemma 16.11, every $C_1' \subseteq C_1$ of size $3m$ has

$$\Delta(G_N[C_1]) \ge \Delta(G_N[C_1']) \ge 2e(G_N[C_1'])/|C_1'| \ge n^2/81.$$

Thus if we take $t_1 = n^{1.99}$ in Lemma 16.1, we find for n sufficiently large that $|C_1| \le 3m$ for all $C_1 \in \mathcal{C}_1$.

We now want to choose t_2 so that each $C_2 \in \mathcal{C}_2(C_1)$ has size very close to m. Let $G' = G_N[C_1]$. If $|C_2| > (1+\gamma)m$ with $\gamma \le 1/3$ we find that

$$\Delta(G'[C_2]) \le 2e(G'[C_2])/|C_2| \le \gamma n/3.$$

With some foresight we take $t_2 = \epsilon n/12$ to guarantee that $|C_2| \le (1+\epsilon/4)m$ for all $C_2 \in \mathcal{C}_2(C_1)$ by Lemma 16.8.

Recall that we want to show with high probability no independent set I of size $(1+\epsilon)pm$ lies in the random set R_p. To this end, we note that we can identify each I with a pair (C_1, S_2) where

- $C_1 \in \mathcal{C}_1$ contains I,
- S_2 is the set $f(I)$ from Lemma 16.3, i.e. $|S_2| \le |C_1|/t_2$, $S_2 \subseteq I$, and S_2 determines a set $C_2 \supseteq I$ in $\mathcal{C}_2(C_1)$.

Given this pair, if $I \subseteq R_p$ has size at least $(1+\epsilon)m$, then (1) $S_2 \subseteq R_p$ since $S_2 \subseteq I$, and (2)

$$|R_p \cap (C_2 \setminus S_2)| \ge (1+\epsilon)pm - |S_2|,$$

since $C_2 \setminus S_2$ contains $I \setminus S_2$. Observe that for $p = c/n$ with $c \gg \epsilon^{-2}$ we have

$$|S_2| \le \frac{|C_1|}{t_1} \approx \frac{m}{\epsilon n} \le \frac{\epsilon pm}{2}.$$

With this in mind, we define $A(S_2)$ to be the event that $S_2 \subseteq R_p$ and $B(S_2)$ to be the event that $|R_p \cap (C_2 \setminus S_2)| \geq (1+\epsilon/2)pm$. Observe that $A(S_2)$ and $B(S_2)$ are independent events, so by a union bound over all pairs (C_1, S_2), we see that the probability that $\alpha(G_N[R_p]) \geq (1+\epsilon)pm$ is at most

$$\sum_{C_1 \in \mathcal{C}_1} \sum_{S_2 : |S_2| \leq |C_1|/t_1} \Pr[A(S_2)] \cdot \Pr[B(S_2)]. \tag{18}$$

We have $\Pr[A(S_2)] = p^{|S_2|}$. By a Chernoff bound it is not difficult to show that $\Pr[B(S_2)] \leq e^{-c'\epsilon^2 pm}$ for some $c' > 0$. Thus if we fix some C_1 in the first term of the sum we get

$$\sum_{s \leq |C_1|/t_2} \binom{|C_1|}{s} p^s \cdot e^{-c'\epsilon^2 pm} \leq \sum_{s \leq |C_1|/t_2} (ep|C_1|/s)^s \cdot e^{-c'\epsilon^2 pm} \approx (p \cdot \epsilon n/12)^{12|C_1|/\epsilon n} \cdot e^{-c'\epsilon^2 pm} \leq (\epsilon c)^{36m/\epsilon n} \cdot e^{-c'\epsilon^2 cm/n},$$

where the approximation only looked at the term with $s = |C_1|/t_2$. Note that for $c \gg \epsilon^{-4}$ the second term dominates, so this bound is roughly

$$e^{-c'\epsilon^2 cm/n}. \tag{19}$$

Note that this is the critical place where we used the two applications of the container lemma: if we only applied the container lemma once to $V(G_N)$ instead of using C_1, then the first term here would be roughly $e^{N/\epsilon n}$ instead of $e^{m/\epsilon n}$, which would dominate the expression.

Returning to (18), we sum the bound of (19) for each element in C_1, which multiplies (19) by

$$\binom{N}{\leq N/t_1} \approx e^{N \log(t_1)/t_1} \approx e^{mn^{-1.49} \log(n)}.$$

This is much smaller than $e^{-c'\epsilon^2 cm/n}$, so the probability in (18) tends to 0 as n tends towards infinity as desired. □

We note that one can get almost as strong a result if one only uses Lemma 16.1, i.e. if one doesn't use the more refined Lemma 16.3. Indeed, the main consequence of using this refined lemma was the extra term $\Pr[A(S_2)] = p^s$ appearing in (18). If one omits this turn, then the same proof will go through provided $p = c \log n/n$. This is a common phenomenon in containers: if you don't use the fact you have certificates $S \subseteq I$, then you'll end up with a bound which is worse by a log factor.

16.3 Counting Sidon Sets

This is a neat example where you use roughly $\log n$ iterated supersaturation lemmas to get the right result. I may write this at some point, but in any case it will be nearly identical to Balogh's notes [12].

17 A Proof of an r-Uniform Container Lemma

In the previous section, we saw how Lemma 16.1 allowed us to effectively count the number of independent sets in "sufficiently nice" graphs. In this section we present a proof of a hypergraph container lemma which applies to "sufficiently nice" hypergraphs, but two things should be noted.

The first is that there are many different variants of hypergraph container lemmas, though most of them are quite similar and broadly speaking apply only to hypergraphs with small codegrees.

The second is that, quite frankly, one doesn't need to know the proof of the container lemma to use it or its variants. As such, the reader may just want to glance at the definition below, and then skip over to latter sections to see some nice applications before jumping back over here whenever they want to see the full proof.

17.1 An Informal Discussion

The hypergraph container lemma we prove comes from [?] (though we deviate somewhat from their notation). We'll formally state this as Theorem 17.5 below, but roughly our goal will be to prove the following. Recall that $\Delta_\ell(H)$ denotes the maximum ℓ-degree of H, i.e. the maximum number of edges containing a given set of ℓ vertices.

Proposition 17.1 (Informal)**.** *If H is an r-graph such that for all $1 \le \ell \le r$ we have*

$$\Delta_\ell(H) \ll q^{\ell-1}\frac{e(H)}{v(H)},$$

then there exists $\delta > 0$, $\mathcal{S} \subseteq \binom{V(H)}{\ll q\cdot v(H)}$, and functions $f : \mathcal{S} \to \binom{V(H)}{\le(1-\delta)v(H)}$ and $g : \mathcal{I}(H) \to \mathcal{S}$ such that

$$g(I) \subseteq I \subseteq f(g(I)) \cup g(I).$$

In other words, if H has small codegrees, then one can find a set of small certificates $\mathcal{S}$ which are each associated with a container $f(S)$ which is of size at most $(1-\delta)v(H)$.

The proof of Proposition 17.1 will in essence be a proof by induction on the uniformity r, and the inductive step of the proof uses an algorithm which is similar to the one used in Lemma 16.1.

Definition 1 (Informal)**.** The *Scythe Algorithm* takes as input a pair (H_{k+1}, I) with H_{k+1} a $(k+1)$-uniform hypergraph and $I \subseteq V(H_{k+1})$ an independent set. It then outputs a triple (H_k, A_k, S_k) with H_k a k-uniform hypergraph such that $I \subseteq V(H_k)$ is an independent set with $I \subseteq A_k \cup S_k$, and S_k is a small set which uniquely determines H_k and A_k given H_{k+1}. Moreover, if H_{k+1} is "nice", then either H_k will be "nice" or A_k will be small.

Given such an algorithm, we can start with any "nice" r-uniform hypergraph H_r and independent set I. We then repeatedly apply this algorithm until we get some H_k which is not "nice", at which point $\bigcup_{i\ge k} S_i$ is a small certificate which determines a small container $A_k \cup \bigcup_{i\ge k} S_i$ for I.

While it's not a priori clear what the "nice" conditions should be, they should in particular guarantee that H_{k+1} has few independent sets, as otherwise there's no hope of this method being effective. In particular, a reasonable set of conditions is to enforce that H_{k+1} has many edges and relatively low codegrees, and this will ultimately be the conditions that we use.

Now that we know what we want our algorithm to do, how should it actually work in practice? Perhaps the most naive approach is to do what we did in Lemma 16.1, where we iteratively select the vertex u of $I \cap H_{k+1}$ which has the largest degree and then adds this to S_k. Once we identify such a u, we know that I does not contain any k-set of the form $e \setminus \{u\}$ for any $e \ni u$ which is an edge in H_{k+1}, so it is natural to make all of these k-sets edges of H_k. If we do this repeatedly, then A_k will be relatively small and I will be an independent set of H_k. Moreover, if we enforce from the start that we'll run this procedure at most s times, then we will have $|S_k| \le s$, giving a small certificate.

Unfortunately we have to be more careful than this. Namely, we need to ensure that H_k is "nice", and in particular that it has small codegrees. As it currently stands this might not work out, e.g. there may be some ℓ-set T which is in many edges containing vertices of S_k. To get around this, we define $D_\ell(H_k, \Delta)$ to be the set of "dangerous" ℓ-sets of $V(H_k)$ which have degree at least $\Delta/2$, where we think of Δ as being the maximum ℓ-degree we want H_k to have (which is analogous to the t parameter used in the algorithm of Lemma 16.1).

We now adjust our naive algorithm by making it so that whenever a set T gets added to $D_\ell(H_k, \Delta)$, we delete from H_{k+1} all of the edges that contain T. This ensures that T never passes over the enforced codegree threshold. With this it turns out that our algorithm will succeed.

17.2 A Formal Algorithm

Motivated by our discussion in the previous section, we make the following definitions. For any hypergraph H', we define the *max-degree order* on $V(H')$ as follows. Fix an arbitrary ordering of $V(H')$. For each integer j, recursively define u_j to be the maximum-degree vertex in the hypergraph $H'[V(H') \setminus \{u_1, \ldots, u_{j-1}\}]$ with ties broken based on the ordering of $V(H')$. The max-degree order is then the ordering $u_1, u_2, \ldots$, and for all j we define $W_{H'}(u_j) = \{u_1, \ldots, u_j\}$.

For any k-uniform hypergraph H', integer $\ell \le k$, and real number Δ, we define

$$D_\ell(H', \Delta) = \left\{ T \in \binom{V(H')}{\ell} : \deg_{H'}(T) \ge \frac{1}{2}\Delta \right\}.$$

Definition 2. The Scythe Algorithm is defined as follows. It takes as input a $(k+1)$-uniform hypergraph H_{k+1} with $k \ge 1$, an independent set $I \subseteq V(H_{k+1})$, and parameters $s, \Delta_1^k, \ldots, \Delta_k^k$.

At the start of the algorithm, we set $H_{k+1}^{(0)} = H_{k+1}$, $S_k^{(0)} = \emptyset$, and we let $H_k^{(0)}$ be the empty hypergraph on $V(H_{k+1})$. For $j = 0, \ldots, s-1$, the algorithm proceeds as follows:

Step 1: If $I \cap V(H_{k+1}^{(j)}) = \emptyset$, then set $H_k = H_k^{(0)}$, $A_k = \emptyset$, $S_k = S_k^{(j)}$. If this happens, stop the algorithm and output (H_k, A_k, S_k).

Step 2: Let u_j be the vertex of $I \cap V(H_{k+1}^{(j)})$ which is first according to the max-degree ordering of $H_{k+1}^{(j)}$. Set $S_k^{(j+1)} = S_k^{(j)} \cup \{u_j\}$.

Step 3: Let $H_k^{(j+1)}$ by the hypergraph on $V(H)$ defined by

$$H_k^{(j+1)} \cup \{e \setminus \{u_j\} : e \in H_{k+1}^{(j)},\ u_j \in e\}.$$

Step 4: Let $H_{k+1}^{(j+1)}$ be the hypergraph on $V(H_{k+1}^{(j)}) \setminus W_{H_{k+1}^{(j)}}(u_j)$ with

$$H_{k+1}^{(j+1)} = \{e \in H_{k+1}^{(j)} : e \cap W_{H_{k+1}^{(j)}}(u_j) = \emptyset \text{ and } T \not\subseteq e \text{ for all } T \in \bigcup_{\ell=1}^{k} D_\ell(H_k^{(j+1)}, \Delta_\ell^k)\}.$$

After running through the above procedure, set $H_k = H_k^{(s)},\ A_k = V(H_{k+1}^{(s)})$, and $S_k = S_k^{(s)}$. Output (H_k, A_k, S_k). ■

We emphasize that this algorithm allows the hypergraphs H_{k+1} and H_k to have repeated edges.

We now analyze this algorithm through a series of lemmas. We omit many of the proofs since most are either straightforward or analogous to what was done in Lemma 16.1.

Lemma 17.2. *Assume one runs the Scythe Algorithm with parameters $s, \Delta_1^k, \ldots, \Delta_k^k$ on inputs $(H_{k+1}, I), (H_{k+1}, I')$ and that the algorithm outputs $(H_k, A_k, S_k), (H_k', A_k', S_k')$, respectively. If $S_k \subseteq I'$ and $S_k' \subseteq I$, then $(H_k, A_k, S_k) = (H_k', A_k', S_k')$.*

For the rest of this subsection we will assume that we have run the Scythe Algorithm with parameters $s, \Delta_1^k, \ldots, \Delta_k^k$ on (H_{k+1}, I) which outputs some (H_k, A_k, S_k). We observe the following basic properties.

Lemma 17.3. *The following hold:*

- *I is an independent set of H_k,*
- *$S_k \subseteq I \subseteq A_k \cup S_k$,*
- *Both H_k and A_k are determined by H_{k+1} and S_k,*
- *$|S_k| \leq s$,*
- *For all $\ell \leq k$ we have $\Delta_\ell(H_k) \leq \frac{1}{2}\Delta_\ell^k + \Delta_{\ell+1}(H_{k+1})$.*

We now turn to the main lemma of this subsection, which roughly says that if H_{k+1} is "nice," then either H_k will also be "nice" or A_k will be small.

Lemma 17.4. *Either*

$$e(H_k) \geq \min\left\{\frac{s}{v(H)}, \frac{\Delta_1^k}{\Delta_1(H_{k+1})}, \ldots, \frac{\Delta_k^k}{\Delta_k(H_{k+1})}\right\} \cdot \frac{e(H_{k+1})}{(k+1)2^{k+2}},$$

or

$$|A_k| \leq v(H_{k+1}) - \frac{e(H_{k+1})}{4 \cdot \Delta_1(H_{k+1})}.$$

Note that this lemma will be most "effective" when each ratio of the minimum is roughly the same. And indeed, we will end up choosing our parameters so that these ratios are all at least $s/v(H)$.

Proof. If the algorithm ever stops at Step 1, then $|A_k| = 0$ and there is nothing to prove, so we can assume that Steps 2 through 4 are completed a total of s times. We observe that

$$|A_k| = v(H_{k+1}^{(s)}) = v(H_{k+1}) - \sum_{j=0}^{s-1} |W_j(u_j)|,$$

where for ease of notation we let $W_j = W_{H_{k+1}^{(j)}}$. Thus we can assume

$$\sum_{j=0}^{s-1} |W_j(u_j)| < \frac{e(H_{k+1})}{4\Delta_1(H_{k+1})}. \tag{20}$$

By construction, for all j we have

$$e(H_k^{(j+1)}) - e(H_k^{(j)}) = \deg_{H_{k+1}^{(j)}}(u_j).$$

Because u_j is the largest element of $I \cap V(H_{k+1}^{(j)})$ in the max-degree order, the degree of u_j is at least as large as the average degree of the subgraph of $H_{k+1}^{(j)}$ after deleting $W_j(u_j) \setminus \{u_j\}$, and again by definition of the max-degree order, this is at least as large as the average degree of $H_{k+1}^{(j)}$ (which has at most $v(H)$ vertices). In total then we find

$$e(H_k) = \sum_{j=0}^{s-1} e(H_k^{(j+1)}) - e(H_k^{(j)}) \geq \sum_{j=0}^{s-1} \frac{(k+1)e(H_{k+1}^{(j+1)})}{v(H)}.$$

If we have $(k+1)e(H_{k+1}^{(j+1)}) \geq e(H_{k+1})$ for all j, then the above sum is at least $(s/v(H)) \cdot e(H_{k+1})$, giving the desired result. Thus we can assume this fails for some j, and this implies

$$e(H_{k+1}^{(s)}) \leq e(H_{k+1}^{(j+1)}) < \frac{e(H_{k+1})}{k+1}. \tag{21}$$

This means that many edges of H_{k+1} were deleted in Step 4 of the algorithm. We claim that this implies that one of the sets $D_\ell(H_k, \Delta_\ell^k)$ is large. Indeed, observe that

$$e(H_{k+1}^{(j)}) - e(H_{k+1}^{(j+1)}) \leq |W_j(u_j)| \cdot \Delta_1(H_k) + \sum_\ell |D_\ell(H_k^{(j+1)}, \Delta_\ell^k) \setminus D_\ell(H_k^{(j)}, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}),$$

since edges are either deleted by deleting vertices in $W_j(u_j)$ or by deleting ℓ-sets which are in D_ℓ for $H_k^{(j+1)}$ but not $H_k^{(j)}$. Summing this over all j gives

$$\begin{aligned} e(H_{k+1}) - e(H_{k+1}^{(s)}) &\leq \sum_j |W_j(u_j)| \cdot \Delta_1(H_{k+1}) + \sum_\ell |D_\ell(H_k, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}) \\ &< \frac{k \cdot e(H_{k+1})}{2(k+1)} + \sum_\ell |D_\ell(H_k, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}), \end{aligned}$$

where this last step used (20) and $k \geq 1$. Using (21) shows that

$$\frac{k \cdot e(H_{k+1})}{2(k+1)} \geq \sum_{\ell} |D_\ell(H_k, \Delta_\ell^k)| \cdot \Delta_\ell(H_{k+1}),$$

so for some ℓ we must have

$$|D_\ell(H_k, \Delta_\ell^k)| \geq \frac{e(H_{k+1})}{2(k+1)\Delta_\ell(H_{k+1})}.$$

With ℓ as above, the handshaking lemma and definition of D_ℓ implies

$$e(H_k) = \binom{k}{\ell}^{-1} \sum_{T \in \binom{V(H_k)}{\ell}} \deg_{H_k}(T) \geq \binom{k}{\ell}^{-1} \cdot |D_\ell(H_k, \Delta_\ell^k)| \cdot \frac{1}{2}\Delta_\ell^k \geq \frac{e(H_{k+1}) \cdot \Delta_\ell^k}{(k+1)2^{k+2}\Delta_\ell(H_{k+1})},$$

giving the desired result. □

17.3 Proof of The Formal Result

We are now ready to prove the following formal statement.

Theorem 17.5 ([14]). *For every integer $r \geq 2$ and $c \geq 1$, there exists $\delta > 0$ such that the following holds. Let $q \in (0,1)$ and suppose H is an r-uniform hypergraph such that for every $1 \leq \ell \leq r$ we have*

$$\Delta_\ell(H) \leq cq^{\ell-1} \cdot \frac{e(H)}{v(H)}.$$

Then there exists $\mathcal{S} \subseteq \binom{V(H)}{\leq (r-1)q \cdot v(H)}$ and functions $f : \mathcal{S} \to \binom{V(H)}{\leq (1-\delta)v(H)}$ and $g : \mathcal{I}(H) \to \mathcal{S}$ such that for every $I \in \mathcal{I}(H)$ we have

$$g(I) \subseteq I \subseteq f(g(I)) \cup g(I).$$

Moreover, $S \cap f(S) = \emptyset$ for all $S \in \mathcal{S}$, and if $I, I' \in \mathcal{I}(H)$ satisfy $g(I) \subseteq I'$, $g(I') \subseteq I$, then $g(I) = g(I')$.

Proof. For all $\ell \leq r$ let $\Delta_\ell^r := \Delta_\ell(H)$, and inductively define

$$\Delta_\ell^k := \max\{2 \cdot \Delta_{\ell+1}^{k+1}, q \cdot \Delta_\ell^{k+1}\}.$$

The following is straightforward to prove given the hypothesis of the theorem.

Claim 17.6. *For all $k < r$ we have $\Delta_1^{k+1} \leq C2^r q^{r-k-1} \frac{e(H)}{v(H)}$.*

For each $I \in \mathcal{I}(H)$, iteratively run through the Scythe Algorithm with the parameters above and $s = q \cdot v(H)$, starting with $H_r = H$. Let $(H_{r-1}, A_{r-1}, S_{r-1}), \ldots, (H_1, A_1, S_1)$ denote the outputs of this algorithm.

It is straightforward to show that $\Delta_\ell(H_k) \leq \Delta_\ell^k$ for all ℓ, k by using induction and Lemma 17.3. For $k < r$ we define $c_k = (Cr2^{r+1})^{k-r}$. Let K be the smallest integer such that $|A_K| \leq$

$(1-c_K)v(H)$, and if no such $K \geq 1$ exists we set $K = 0$. It is straightforward to prove that $e(H_k) \geq c_k q^{k-r} e(H)$ for all $k > K$ by using induction, Lemma 17.4, and Claim 17.6.

Let $\delta := c_1$. Before we define our functions, for technical reasons it will be convenient to first define a function $f^* : \mathcal{I}(H) \to \binom{V(H)}{\leq (1-\delta)v(H)}$ before defining f. Pick some I and let the hypergraphs H_k and integer K be as defined above. Observe that if $K \geq 1$, then $|A_K| \leq (1-\delta)v(H)$, so in this case we will set

$$g(I) = \bigcup_{k \geq K} S_k, \quad f^*(I) = A_K.$$

If $K = 0$, then we set

$$g(I) = \bigcup_{k \geq 1} S_k, \quad f^*(I) = \{v \in V(H_1) : \{v\} \notin H_1\}.$$

Note that in this case $|f^*(I)| = v(H) - e(H_1)$, which is at most $(1-\delta)v(H)$ by our observations above. Lastly, we define $\mathcal{S} = \{g(I) : I \in \mathcal{I}(H)\}$ and $f(S) = f^*(I)$ for any $I \in g^{-1}(S)$. The fact that f is well defined is implied by the following claim, which itself follows from Lemma 17.3.

Claim 17.7. *If $I, I' \in \mathcal{I}(H)$ with $g(I) \subseteq I'$ and $g(I') \subseteq I$, then $g(I) = g(I')$ and $f^*(I) = f^*(I')$.*

The fact that these definitions give the desired result, except possibly the condition $S \cap f(S) = \emptyset$, can be checked by using the properties from Lemma 17.3. This last condition can be established by taking $f'(S) := f(S) \setminus S$ if needed. $\square$

The following weaker version of Theorem 17.5 is often good enough for most applications and is conceptually simpler.

Corollary 17.8. *For every integer $r \geq 2$ and $c \geq 1$, there exists $\delta > 0$ such that the following holds. Let $q \in (0, 1)$ and suppose H is an r-uniform hypergraph such that for every $1 \leq \ell \leq r$ we have*

$$\Delta_\ell(H) \leq c q^{\ell-1} \cdot \frac{e(H)}{v(H)}.$$

Then there exists a collection of sets $\mathcal{C}$ such that every independent set of H is a subset of some $C \in \mathcal{C}$, and moreover, $|C| \leq (1-\delta)v(H)$ for all $C \in \mathcal{C}$ and $|\mathcal{C}| \leq \binom{v(H)}{\leq (r-1)q \cdot v(H)}$.

Proof. In the notation of Theorem 17.5, we let $\mathcal{C} = \{f(g(I)) \cup g(I) : I \in \mathcal{I}(H)\}$. $\square$

On its own, Corollary 17.8 (and even Theorem 17.5) isn't terribly useful since the containers it generates are rather large, and in practice one needs to reapply this lemma to each $C \in \mathcal{C}$ which is large, and to keep repeating this argument until the contains are sufficiently small. Because δ depends only on r and C, this only needs to be done a constant number of times. However, to reapply the lemma, each large $C \in \mathcal{C}$ must satisfy essentially the same hypothesis as H. While a generic hypergraph will fail to have this property, many nice hypergraphs will.

18 Hypergraph Containers and Triangle-Free Graphs

Let us restate our weak container theorem Corollary 17.8 for the special case of 3-uniform hypergraphs.

Theorem 18.1 ([14])**.** *For every $c \geq 1$, there exists $\delta > 0$ such that the following holds. Let $q \in (0,1)$ and suppose H is a 3-uniform hypergraph such that*

$$\Delta_1(H) \leq c\frac{e(H)}{v(H)},$$

$$\Delta_2(H) \leq cq\frac{e(H)}{v(H)},$$

$$\Delta_3(H) \leq cq^2\frac{e(H)}{v(H)}.$$

Then there exists a collection of sets $\mathcal{C}$ such that every independent set of H is a subset of some $C \in \mathcal{C}$, and moreover, $|C| \leq (1-\delta)v(H)$ for all $C \in \mathcal{C}$ and $|\mathcal{C}| \leq \binom{v(H)}{\leq 2q\cdot v(H)}$.

Note that for simple hypergraphs we have $\Delta_3(H) = 1$, so this last bound is equivalent to lower bounding the average degree by $c^{-1}q^{-2}$. One important consequence of Theorem 18.1 is the following.

Theorem 18.2. *For all $n, \epsilon > 0$, there exists a collection of n-vertex graphs $\mathcal{C}$ such that*

(a) Every triangle-free graph $G \subseteq K_n$ is a subgraph of some $C \in \mathcal{C}$,

(b) Every $C \in \mathcal{C}$ has less than ϵn^3 triangles, and

(c) We have $|\mathcal{C}| = n^{O_\epsilon(n^{3/2})}$.

That is, there exists a small set of nearly triangle-free graphs which contains every triangle-free graph.

Proof. Start with $\mathcal{C} = \{K_n\}$, and note that $\mathcal{C}$ trivially satisfies (a). Iteratively proceed as follows. If every $C \in \mathcal{C}$ has less than ϵn^3 triangles then output the current collection $\mathcal{C}$. Otherwise, let $C \in \mathcal{C}$ be such that it contains at least ϵn^3 triangles. Form a 3-graph H with vertex set $E(C)$ where three edges of C form a hyperedge in H if they form a triangle. Note that $e(H) \geq \epsilon n^3$ and $v(H) = e(C) \leq n^2$. Every edge is contained in at most n triangles, so $\Delta_1(H) \leq n \leq \epsilon^{-1}\frac{e(H)}{v(H)}$. We also have $\Delta_2(H) = \Delta_3(H) = 1 \leq \epsilon^{-1}(n^{-1/2})^2\frac{e(H)}{v(H)}$. With this we see that we can apply Theorem 18.1 with $q = n^{-1/2}$ and $c = \epsilon^{-1}$. This gives a collection of containers $\mathcal{C}'$ for C, i.e. subgraphs $C' \subseteq C$ such that every triangle-free subgraph of C is contained in some $C' \in \mathcal{C}$. Remove C from $\mathcal{C}$ and add every $C' \in \mathcal{C}'$ to $\mathcal{C}$. Repeat this process.

Let $\mathcal{C}$ be the final collection that this algorithm produces. It is straightforward to show that (a) holds inductively, and (b) holds by construction. To show that the final collection is small, first note that each time we apply the container lemma, the number of new graphs we create is at most $\binom{v(H)}{\le 2n^{-1/2}v(H)} = n^{O(n^{3/2})}$. Second, observe that each time we apply the container lemma to C, the graphs in $\mathcal{C}'$ have at most $(1-\delta)e(C)$ edges, where δ depends only on ϵ. Because we only iterate on C which have at least ϵn^2 edges (since they need at least ϵn^3 triangles), we iteratively apply the lemma at most some bounded number of times $b = b(\epsilon)$ to reach any element in the final collection $\mathcal{C}$. Thus the total number of containers we create is $\left(n^{O(n^{3/2})}\right)^b = n^{O_\epsilon(n^{3/2})}$ as desired. ☐

For Theorem 18.2 to be useful, we need to get a handle on graphs with at most ϵn^3 triangles. As is typical with containers, this will come from a supersaturation lemma.

Lemma 18.3. *For every $\delta > 0$ there exists an $\epsilon > 0$ such that if G is an n-vertex graph with $e(G) \ge (\frac{1}{2} + \delta)\binom{n}{2}$, then G contains at least ϵn^3 triangles.*

I'm not crazy about the ordering of δ, ϵ but I admit the final thing should be about ϵ...well actually in a lot of the applications it kind of makes more sense to do it the other way. Maybe do the K_r version in general depending on what I need later on.

Proof. TODO ☐

With this we can prove the following counting result.

Theorem 18.4. *The number of n-vertex triangle-free graphs is equal to*

$$2^{(1+o(1))n^2/4}.$$

Proof. The lower bound comes from considering all of the subgraphs of $K_{n/2,n/2}$. For the upper bound, fix some $\delta > 0$ and let ϵ be as in Lemma 18.3. Let $\mathcal{C}$ be the containers guaranteed by Theorem 18.2 with parameter ϵ. Because every triangle-free graph is a subgraph of some $C \in \mathcal{C}$, the number of triangle-free graphs is at most

$$\sum_{C\in\mathcal{C}} 2^{|C|} \le n^{O(n^{3/2})} \cdot 2^{\max_{C\in\mathcal{C}} e(C)},$$

Since each $C \in \mathcal{C}$ has less than ϵn^3 and at triangles, Lemma 18.3 implies $e(C) \le (\frac{1}{2} + \delta)\binom{n}{2}$ for all $C \in \mathcal{C}$. In total we get an upper bound of

$$2^{(\frac{1}{2}+\delta)\binom{n}{2}+O(n^{3/2}\log n)},$$

and letting δ tend towards 0 gives the result. ☐

While containers most directly allow one to solve problems that are equivalent to counting the number of independent sets of a hypergraph, there are other related problems which they're effective for. For example, in the next subsection we show how containers can be used to prove probabilistic analogs of classical extremal results. In the section after this we show how one can use containers to count special kinds of independent sets, namely maximal independent sets.

18.1 Mantel's Theorem in Random Graphs

Given two graphs G, F, we let $\mathrm{ex}(G, F)$ denote the largest F-free subgraph of G. For example, $\mathrm{ex}(K_n, F) = \mathrm{ex}(n, F)$. The following result can be viewed as a random version of Mantel's theorem.

Theorem 18.5. *Define* $\mathrm{ex}(G_{n,p}, K_3)$ *to be the largest triangle-free subgraph of* $G_{n,p}$. *We have* $\mathrm{ex}(G_{n,p}, K_3) = (1 + o(1))pn^2/4$ *whp provided* $p \gg n^{-1/2} \log n$.

Proof. The lower bound follows by considering $G_{n,p} \cap K_{n/2,n/2}$, which is always triangle-free and which has $(1 + o(1))pn^2/4$ edges whp. For the upper bound, fix $\delta > 0$, and let $\epsilon > 0$ be as in Lemma 18.3. Let $\mathcal{C}$ be the set of containers given by Theorem 18.2 with parameter ϵ, and as before we have $e(C) \le (1/2+\delta)\binom{n}{2}$ for all $C \in \mathcal{C}$. Because every triangle-free graph is contained in some $C \in \mathcal{C}$, in order to have $\mathrm{ex}(G_{n,p}, K_3) \ge (1 + 4\delta)pn^2/4$, there must exist some $C \in \mathcal{C}$ such that $|G_{n,p} \cap C| \ge (1+4\delta)pn^2/4$. Let E_C be the event that this bound holds. Observe that $|G_{n,p} \cap C|$ is a binomial random variable with probability p and at most $(1+2\delta)n^2/4$ trials. By the Chernoff bound, we find $\Pr[E_C] \le e^{-O_\delta(pn^2)}$. In total then, we have

$$\Pr[\mathrm{ex}(G_{n,p}, K_3) \ge (1 + 4\delta)pn^2/4] \le \Pr\left[\bigcup_{C \in \mathcal{C}} E_C\right] \le n^{O_\delta(n^{3/2})} \cdot e^{-O_\delta(pn^2)} \to 0,$$

with this last step holding by hypothesis on p. We conclude the reuslt by taking δ arbitrarily close to 0. $\square$

We note that for $p \ll n^{-1/2}$, a simple deletion argument shows that for $p \ll n^{-1/2}$ there exist triangle-free subgraphs with $(1 + o(1))p\binom{n}{2}$ edges, and this is certainly best possible since $G_{n,p}$ has at most this many edges asymptotically. Thus the bound for p in Theorem 18.5 is almost optimal. In fact, we can obtain the optimal bound in this theorem by using the strong container theorem Theorem 17.5, which in the case of 3-graphs can be written as follows.

Theorem 18.6. *For every* $c \ge 1$, *there exists* $\delta > 0$ *such that the following holds. Let* $q \in (0, 1)$ *and suppose* H *is a* 3*-uniform hypergraph such that*

$$\Delta_1(H) \le c\frac{e(H)}{v(H)},$$

$$\Delta_2(H) \le cq\frac{e(H)}{v(H)},$$

$$\Delta_3(H) \le cq^2\frac{e(H)}{v(H)}.$$

Then there exists $\mathcal{S} \subseteq \binom{V(H)}{\le 2q \cdot v(H)}$ *and functions* $f : \mathcal{S} \to \binom{V(H)}{\le (1-\delta)v(H)}$ *and* $g : \mathcal{I}(H) \to \mathcal{S}$ *such that for every* $I \in \mathcal{I}(H)$ *we have*

$$g(I) \subseteq I \subseteq f(g(I)) \cup g(I).$$

Moreover, $S \cap f(S) = \emptyset$ *for all* $S \in \mathcal{S}$, *and if* $I, I' \in \mathcal{I}(H)$ *satisfy* $g(I) \subseteq I'$, $g(I') \subseteq I$, *then* $g(I) = g(I')$.

This allows us to construct the following "strong" set of containers for triangle-free graphs.

Theorem 18.7. *Let $\mathcal{G}_n, \mathcal{T}_n$ denote the set of all n-vertex graphs and all n-vertex triangle-free graphs, respectively. For all $n, \epsilon > 0$, there exists a set of graphs $\mathcal{S}$ with at most $O_\epsilon(n^{3/2})$ edges, as well as functions $f : \mathcal{S} \to \mathcal{G}_n$ and $g : \mathcal{T}_n \to \mathcal{S}$ such that for every $G \in \mathcal{T}_n$, we have*

$$g(G) \subseteq G \subseteq f(g(G)) \cup g(G),$$

and such that $f(S)$ has less than ϵn^3 triangles for all $S \in \mathcal{S}$.

Proof. We start with $\mathcal{S}$ consisting only of the empty graph and define $g(G) = \emptyset$ and $f(\emptyset) = K_n$. Iteratively assume we have constructed some $\mathcal{S}, f, g$ satisfying all of the conditions except possibly that each $S \in \mathcal{S}$ has at most $O_\epsilon(n^{3/2})$ edges and that $f(S)$ has less than ϵn^3 triangles (which holds for our initial step). If $f(S)$ has less than ϵn^3 triangles for all $S \in \mathcal{S}$ then we end the procedure. Otherwise, let S be such that $C = f(S)$ has at least ϵn^3 triangles. By repeating our computations from the proof of Theorem 18.2, we see that we can apply Theorem 18.6 to the 3-graph H encoding triangles of C, and we let $\mathcal{S}_C, f_C, g_C$ be the output of this theorem.

Claim 18.8. *Let $\mathcal{S}' := (\mathcal{S} \setminus \{S\}) \cup \{S_C \cup S : S_C \in \mathcal{S}_C\}$, define $g'(G) = g(G)$ if $g(G) \neq S$ and $g'(G) = g_C(G - S)$ otherwise, and define $f'(S') = f(S')$ if $S' \in \mathcal{S} \setminus \{S\}$ and $f'(S') = f_C(S' - S)$ otherwise. These maps are well defined and satisfy the conditions of the theorem except possibly that each $S \in \mathcal{S}'$ has at most $O_\epsilon(n^{3/2})$ edges and that $f'(S)$ has less than ϵn^3 triangles.*

Proof. First observe that because $C \cap S = \emptyset$, each element of $\mathcal{S}_C$ (which is a subgraph of C) is disjoint from S. This implies that all of the elements $S_C \cup S$ for $S_C \in \mathcal{S}_C$ are distinct. Moreover, none of these elements are equal to any element of $\mathcal{S} \setminus \{S\}$. Indeed, if $S_C \cup S = S' \in \mathcal{S}$, then $\mathcal{S}$ would contain two elements with $S \subsetneq S'$. The last condition of Theorem 18.6 then implies that we must have $S = S'$. This all implies that g', f' are well defined maps, and it is not difficult to check that they inherit all of the other desired properties. □

With this we can keep applying Theorem 18.6 until we get $\mathcal{S}, f, g$ which satisfies all of the conditions except possibly that $e(S)$ is small. As in the proof of Theorem 18.2, one can check that each $S \in \mathcal{S}$ is obtained by applying Theorem 18.6 at most $O_\epsilon(1)$ times, and each time its applied at most $O(n^{3/2})$ edges get added to S. With this we can conclude the result. □

We note that there exists a somewhat stronger version of Theorem 18.6 (and more generally Theorem 17.5) which allows one to prove the previous result with less work. However, the theorem statement is somewhat more complicated conceptually (involving things called $(\mathcal{F}, \epsilon)$-dense families), so for this exposition we have opted to use the simpler version. In any case, with this enhanced version of Theorem 18.2, we can improve upon our threshold for the random Mantel theorem by dropping a logarithmic term.

Theorem 18.9. *Define $\mathrm{ex}(G_{n,p}, K_3)$ to be the largest triangle-free subgraph of $G_{n,p}$. We have $\mathrm{ex}(G_{n,p}, K_3) = (1 + o(1))pn^2/4$ whp provided $p \gg n^{-1/2}$.*

Proof. The lower bound follows by considering $G_{n,p} \cap K_{n/2,n/2}$, which is always triangle-free and which has $(1 + o(1))pn^2/4$ edges whp. For the upper bound, fix $\delta > 0$, and let $\epsilon > 0$

be as in Lemma 18.3. Let $\mathcal{S}, f, g$ be as in Theorem 18.7. Note that each $f(S)$ has at most $(1/4+2\delta)n^2$ edges by Lemma 18.3. For each $S \in \mathcal{S}$, let E_S be the event that $S \subseteq G_{n,p}$ and that $|f(S) \cap G_{n,p}| \geq (1+4\delta)pn^2/4$. Note that in order to have $\mathrm{ex}(G_{n,p}, K_3) \geq (1+4\delta)pn^2 + O_\epsilon(n^{3/2})$, some E_S event must occur, and moreover that $\Pr[E_S] = p^{|S|} \cdot e^{-O_\delta(pn^2)}$. With this we have

$$\Pr[\mathrm{ex}(G_{n,p}, K_3) \geq (1+4\delta)pn^2/4 + O_\epsilon(n^{3/2})] \leq \Pr\left[\bigcup_{S \in \mathcal{S}} E_S\right] \leq \sum_{s=0}^{O_\epsilon(n^{3/2})} \sum_{S \in \mathcal{S}: |S|=s} p^s e^{-O_\delta(pn^2)}.$$

As the number of $S \in \mathcal{S}$ with $|S| = s$ is trivially at most $\binom{n^2}{s} \leq (en^2/s)^s$, we find that the above is at most

$$\sum_{s=0}^{O_\epsilon(n^{3/2})} (epn^2/s)^s e^{-O_\delta(pn^2)}.$$

One can check that the function $(epn^2/s)^s$ is increasing for $s \leq pn^2$. Since we know $s \leq C_\epsilon n^{3/2}$ for some suitable C_ϵ, we get that the sum above is at most

$$C_\epsilon n^{3/2} \cdot (eC_\epsilon^{-1} pn^{1/2})^{C_\epsilon n^{3/2}} e^{-O_\delta(pn^2)},$$

and this tends to 0 provided $pn^{1/2} \to \infty$ (since $pn^2 \gg n^{3/2} \log(pn^{1/2})$), proving the result. $\square$

Note that in this proof, the main extra power we gained by utilizing Theorem 18.7 is that S must be contained in our subgraph. This makes it so that the $S \in \mathcal{S}$ with many edges "cost more", allowing us to gain.

We note that in general, it is very common that by using the weak container lemma, one ends up getting tight bounds up to a logarithmic factor, and this extra factor can usually be remedied by utilizing the strong container lemma in some straightforward (if slightly more tedious) way.

18.2 Maximal Triangle-Free Graphs

In this subsection we use containers to count maximal independent sets, i.e. those that are maximal with respect to set inclusion. To do this, we again apply the container lemma to find a small collection of containers $\mathcal{C}$ for a hypergraph H. We then argue that each $C \in \mathcal{C}$ contains few maximal independent sets, which gives the result.

In order for this approach to be effective, we need a supersaturation result saying that if $H[C]$ has few edges, then C contains few maximal independent sets, and these results are typically a bit more complicated to prove compared to the non-maximal setting, and often these proofs invoke facts about the number of maximal independent sets in special kinds of graphs.

One case where we can pull off the scheme outlined above is in counting maximal triangle-free graphs. By Theorem 18.4, we know that there are $2^{n^2/4+o(n^2)}$ triangle-free graphs on n vertices, and moreover, it is a well known result of Erdős, Kleitman, and Rothschild [60] that almost all of these graphs are bipartite. However, a bipartite triangle-free graph is a maximal triangle-free graph if and only if it is a complete bipartite graph, and there are less than 2^n such graphs. A different set of constructions gives the following.

Proposition 18.10. *There are at least $2^{n^2/8}$ maximal triangle-free graphs on n-vertices.*

Proof. Write the vertices of $[n]$ as $a_1, \ldots, a_{n/4}, b_1, \ldots, b_{n/4}, c_1, \ldots, c_{n/2}$. Add every edge of the form $a_i b_i$. For every $1 \le i \le n/4$ and $1 \le j \le n/2$, add exactly one of the edges $a_i c_j$ or $b_i c_j$. In total this gives $2^{(n/4)(n/2)} = 2^{n^2/8}$ different graphs $\{G_1, G_2, \ldots, \}$, and it's not hard to see that each of these are triangle-free. Let G_k' be any maximal triangle-free graph containing G_k (so $G_k' = G_k$ if G_k is maximal). One can check that each of the G_k' graphs are distinct from each other, giving the desired result. $\square$

Our goal is to show that this result is best possible. This was originally proven by Balogh and Petříčková [16], and our proof follows their same approach.

Similar to the proof of Theorem 18.4 where we counted the number of triangle-free graphs, we'll begin by constructing a set of using Theorem 18.2 to obtain a set of containers $\mathcal{C}$ where each element is a graph which contains few triangles. From there it remains to show that each of these $C \in \mathcal{C}$ contains at most roughly $2^{n^2/8}$ maximal triangle-free graphs.

Essentially the only thing we know about each $C \in \mathcal{C}$ is that they contain few triangles, and fortunately quite a lot can be said about such graphs: the triangle removal lemma says that we can delete a small number of edges from C to get a triangle-free graph, and supersaturation says that $e(C)$ is not much larger than $\frac{1}{4}n^2$. With these ideas in mind we prove the following.

Lemma 18.11. *Let $\epsilon, \gamma > 0$ be constants. Let $\delta(\epsilon)$ be as in* ***REF****, $\delta(\gamma)$ be as in* ***REF****, and $\delta = \min\{\delta(\epsilon), \delta(\gamma)\}$. If G is an n-vertex graph with at most δn^3 triangles, then the number of maximal triangle-free subgraphs of G is at most*

$$2^{n^2/8 + \gamma n^2/2 + \epsilon n^2}.$$

Proof. By **Triangle-removal**, there exists a set $F \subseteq E(G)$ of size at most ϵn^2 such that $G - F$ is triangle-free. For $F' \subseteq F$, let $\mathcal{M}(F')$ denote the set of maximal triangle-free subgraphs $G' \subseteq G$ with $G' \cap F = F'$. Observe that the $\mathcal{M}(F')$ sets partition the maximal triangle-free subgraphs into $2^{\epsilon n^2}$ sets, so it suffices to show that for any $F' \subseteq F$,

$$|\mathcal{M}(F')| \le 2^{n^2/8 + \gamma n^2/2}.$$

The result is trivial if F' contains a triangle (since there are 0 graphs which contain F' and which are triangle-free), so from now on we'll assume F' is triangle-free. Define an auxiliary graph T via

$$V(T) = G - (F - F') - \{e : \exists f, g \in F' \text{ and } e, f, g \text{ form a triangle}\},$$

and where two edges $f, g \in V(T)$ are adjacent in T if there exists an edge $e \in F'$ which forms a triangle with f, g. We make these definitions so that the following holds.

Claim 18.12. *Every $G' \in \mathcal{M}(F')$ is a maximal independent set of T.*

Proof. Fix some $G' \in \mathcal{M}(F')$. We first claim that

$$G' \subseteq V(T) = G - (F - F') - \{e : \exists f, g \in F' \text{ and } e, f, g \text{ form a triangle}\}.$$

Indeed, we must have $G' \subseteq G - (F - F')$ in order to have $G' \cap F = F'$, and given this, G' cannot contain any edge e for which there exist $f, g \in F'$ forming a triangle with e since G' is triangle-free. This proves the claim.

We next claim that G' is an independent set of T. Indeed, if $f, g \in G'$ were adjacent in T, then there exists some edge $e \in F' \subseteq G'$ forming a triangle with f, g, contradicting G' being triangle-free.

It remains to show that G' is a maximal independent set. To this end, consider any $f \in V(T) \setminus G'$. Because G' is a maximal triangle-free graph, there must exist edges $e, g \in G'$ which form a triangle with f. Observe that at least one of e, g must be in F', since by assumption of $G - F$ being triangle-free, every triangle must contain at least one edge of F (and the only such edges in $V(T)$ are in F'). Thus we may assume $e \in F'$. We also note that g does not form a triangle with two edges of F', as this would contradict $g \in G'$ and G' being triangle-free. Thus we have $f, g \in V(T)$, and $e \in F'$ implies that these two edges are adjacent in T. This implies that G' is indeed a maximal independent of T, completing the proof. $\square$

It remains to show that T has few maximal independent sets. For this we make the following key observation.

Claim 18.13. *The graph T is triangle-free.*

Proof. Assume for contradiction that there existed edges f_1, f_2, f_3 forming a triangle in T. By definition of T, this is only possible if these edges are intersecting, and hence these edges either form a triangle or a star. If they form a triangle, then by definition of T we in fact have $f_i \in F'$ for all i, and hence F' forms a triangle, a contradiction to our assumption on F'. If these edges form a star, say with leaves x, y, z, then again by definition of T we must have $xy, yz, xz \in F'$, a contradiction to our assumption of F' being triangle-free. We conclude the result. $\square$

Thus we've reduced the problem to upper bounding the number of maximal independent sets in a triangle-free graph T. A well known result of Moon and Moser [119] says that an N-vertex triangle-free graph has at most $2^{N/2}$ maximal independent sets, so the number of maximal independent sets of T is at most

$$2^{|V(T)|/2} \leq 2^{e(G)/2} \leq 2^{n^2/8+\gamma n^2/2},$$

where the last step used $\delta \leq \delta(\gamma)$. This implies the result. $\square$

From here one can prove that there are at most $2^{n^2/8+o(n^2)}$ maximal triangle-free graphs on n-vertices analogous to how we proved Theorem 18.4 I should check this more carefully.

Before moving on, it is natural to ask how many maximal K_4-free graphs there are. Unfortunately the present argument completely fails to generalize to this setting.

Essentially, the issue is that T can be viewed as the link graph of a 3-uniform hypergraph. More precisely, let H be the 3-graph with $V(H) = G - (F - F')$ and whose hyperedges are triangles. Then the link hypergraph $H[F']$ contains the graph T together with a 1-edge on every edge e such that there exist $f, g \in F'$ forming a triangle with e. For the K_4 problem, one could again consider the link set of some 6-graph, but the structure you end up getting is some very non-uniform hypergraph which avoids some strange set of subgraphs, and as such the analysis becomes somewhat unwieldy.

19 TODOCounting F-free graphs

Ferber-McKinley-Samotij.

Part VI

Matchings in Hypergraphs

20 Pippenger's Theorem and the Rödl Nibble

Throughout this part we recall that for a hypergraph H and a set of vertices S, we let $\deg_H(S)$ denote the number of edges of H containing S and define $\Delta_i(H) = \max_{S \subseteq \binom{V(H)}{i}} \deg_H(S)$.

We saw in our previous part on hypergraph containers that many problems in extremal combinatorics can be rephrased in terms of independent sets of an appropriate auxiliary hypergraph. In this part we similarly observe the same phenomenon for *matchings* in hypergraphs rather than independent sets, where we recall that a matching M is a set of pairwise-disjoint edges of a hypergraph H.

In fact, studying matchings in hypergraphs is just a (very) special case of studying independent sets in graphs: given a hypergraph H, we can define its *line graph* $L(H)$ to be the graph with vertex set $E(H)$ where two $e, f \in H$ are adjacent in $L(H)$ if and only if $e \cap f \neq \emptyset$; in which case matchings in H are exactly independent sets in $L(H)$.

Because matchings are a very special kind of independent set, one might hope that more can be said about them compared to general independent sets, and this will indeed turn out to be the case. In particular, we will study a number of powerful theorems themed around finding (almost) perfect matchings in "nice" hypergraphs. The first of these results is the following fundamental theorem of Pippenger **REF**(building upon earlier results of Rödl [130]), which says that nearly regular hypergraphs with small codegrees have almost perfect matchings. Here we write $c = 1 \pm \delta$ as shorthand for saying that c is a real number in the interval $[1-\delta, 1+\delta]$.

Theorem 20.1 (Pippenger's Theorem). *For every $r \geq 2$ and reals $K \geq 1$ and $a > 0$, there are $\delta = \delta(r, K, a) > 0$ and $D_0 = D_0(r, K, a)$ such that the following holds for every $D \geq D_0$: let H be an r-graph such that*

(i) For all but at most $\delta v(H)$ vertices $x \in V$, we have $\deg(x) = (1 \pm \delta)D$,

(ii) $\Delta_1(H) < KD$, and

(iii) $\Delta_2(H) < \delta D$.

Then there exists a matching of H using at least $(1-a)(v(H)/r)$ edges.

While the proof of Pippenger's Theorem uses only elementary tools and contains a number of important ideas, the full details of the argument are rather annoying to write down (see for example the proof detailed in [8, Theorem 4.7.1] which uses over 20 named constants!). As such, we will omit going through this complete proof here, instead electing to focus on some applications of the result followed by a high-level look at the proof ideas.

20.1 Applications

Historically, the first application of (the precursor to) Pippenger's Theorem was done by Rödl in order to solve a problem in design theory. To state this, we say that a hypergraph $\mathcal{S}$ is a *partial (n,q,k)-Steiner system* if $\mathcal{S}$ is an n-vertex q-uniform hypergraph such that every k-set of vertices is contained in at most one edge, and we say that $\mathcal{S}$ is a *(n,q,k)-Steiner system* if we further have that every k-set is contained in exactly one edge.

It is not difficult to see that every partial (n,q,k)-Steiner system has at most $\binom{q}{k}^{-1}\binom{n}{k}$ edges, with equality holding if and only if $\mathcal{S}$ is an (n,q,k)-Steiner system. Erdös and Hannini [59] asked whether one could find partial Steiner systems with asymptotically this many edges, and this was answered positively by Rödl.

Theorem 20.2 (Rödl [130])**.** *For all fixed $q \geq k$, there exists a partial (n,q,k)-Steiner system $\mathcal{S}$ with*

$$|\mathcal{S}| \geq (1-o(1))\binom{q}{k}^{-1}\binom{n}{k}.$$

Proof. In order to apply Pippenger's Theorem, we need to construct a hypergraph where large matchings correspond to large partial (n,q,k)-Steiner systems. To this end, we define H to be the $r := \binom{q}{k}$-uniform hypergraph whose vertex set is $\binom{[n]}{k}$ and whose edge set is $\{\binom{Q}{k} : Q \in \binom{[n]}{q}\}$. That is, each hyperedge of H consists of all the sets of size k that are covered by a given edge $Q \in E(H_n^q)$.

It is not too difficult to check that Theorem 20.2 is equivalent to saying that there exists a matching of H of the stated size. Moreover, it is easy to check that H is $\binom{n-k}{q-k}$-regular and that every pair of vertices is in at most $\binom{n-k-1}{q-k-1} \ll \binom{n-k}{q-k}$ edges. We can thus apply Pippenger's Theorem with $v(H) = \binom{n}{k}$, $r = \binom{q}{k}$, and $D = \binom{n-k}{q-k}$ to give a matching of the desired size. $\square$

Theorem 20.2 shows that "approximate" Steiner systems exist, and it is natural to ask when Steiner systems exist. The simplest non-trivial case is $S(n,3,2)$, which are also known as Steiner triple systems. It is not difficult to see that if a Steiner triple system on n vertices exists, then $3|\binom{n}{2}$ (each edge covers 3 pairs and each of the $\binom{n}{2}$ pairs are covered exactly once) and $2|(n-1)$ (for any given vertex v, each edge contains 2 pairs containing v and there are exactly $n-1$ such pairs). Equivalently, this argument says that if a Steiner triple system exists, then it is necessary that $n \equiv 1,3 \mod 6$. It turns out that this condition is also sufficient due to certain constructions involving quasigropus and latin squares.

In general for an $S(n,q,k)$ to exist, there are certain "obvious" divisibility conditions that must be satisfied, but in general these are not sufficient. In fact, as of 2014, it wasn't even known if, say, any $S(n,q,6)$ Steiner systems existed, let alone if there were infinitely many n for which such a Steiner system existed. In a major breakthrough, it was shown by Keevash [98] and independently by Glock, Kühn, Lo, and Osthus [80] that if n is sufficiently large in terms of q,k, then $S(n,q,k)$ systems exist if and only if n satisfies the obvious divisibility conditions. The core of Keevash's proof was a variant of the proof technique we discuss below for proving Pippenger's Theorem done in a more algebraic setting, but the proof is very, very complicated!

Todo: insert Glock's application to Brown-Erdős-Sós.

20.2 The Rödl Nibble

Informally, Pippenger's Theorem says that every "nice" r-graphs H contains a matching covering almost all of its vertices. At a very high-level, our approach for constructing such a matching will go as follows:

- We start by randomly choosing $\epsilon n/r$ edges E_1 from H for some small $\epsilon > 0$. With high probability E_1 will cover about $e^{-\epsilon}n$ vertices.
- We delete vertices covered by E_1 to get a new hypergraph H_2. With high probability H_2 will also be "nice", so conditional on this we can iterate the procedure above and pick some random set of edges E_2 from H_2 to cover some more vertices.
- We keep doing this until ϵn vertices remain uncovered, at which point we stop the process.

The approach outlined above is broadly known as the *Rödl nibble* or the *semirandom method*. To reiterate, the core idea is that you iteratively do something to a small chunk of vertices in such a way that the structure of the rest of your hypergraph remains the same with high probability, which allows one to continuously iterate this procedure until one is left with a very small leftover part. For our purposes, we will specifically want to prove the following.

Lemma 20.3. *For every $r \geq 2$ and reals $K \geq 1$ and $\epsilon, \delta' > 0$, there are $\delta = \delta(r, K, \epsilon, \delta') > 0$ and $D_0 = D_0(r, K, \epsilon, \delta')$ such that for every $n \geq D \geq D_0$ the following holds.*

Let $H = (V, E)$ be an n-vertex r-graph such that

(i) For all but at most δn vertices $x \in V$, we have $\deg(x) = (1 \pm \delta)D$,

(ii) $\Delta_1(H) < KD$, and

(iii) $\Delta_2(H) < \delta D$.

In this case there exist a set of edges $E' \subseteq E$ such that

(a) $|E'| = (1 \pm \delta')(\epsilon n/r)$,

(b) The set $V' := V - \bigcup_{e \in E'} e$ has $|V'| = (1 \pm \delta')e^{-\epsilon}n$, and

(c) For all but at most $\delta'|V'|$ vertices $x \in V'$, the degree $\deg'(x)$ of x in the induced hypergraph $H[V']$ satisfies $\deg'(x) = (1 \pm \delta')De^{-\epsilon(r-1)}$.

The main point of the conclusion here is that the number of uncovered vertices and their degrees shrink in a predictable way, allowing one to repeatedly apply this result with carefully chosen values of ϵ, δ to prove Pippenger's Theorem. Again, we omit the exact details of this reduction and only give a high-level sketch of how to prove this key lemma.

Sketch of Proof. Throughout the proof we'll introduce various constants δ_i which we assume to be sufficiently small in terms of our relevant parameters. Our proof will only make use of Chebyshev's inequality in order to give a thorough example of implementing the second moment method in practice, though we emphasize that e.g. the Chernoff bound could alternatively be used in various places.

Randomly choose a subset $E' \subseteq E$ such that each edge of E appears in E' independently and with probability $p := \epsilon/D$. Roughly speaking, our goal will be to show that with this choice of E', each of (a),(b),(c) occur in expectation, after which we will use Chebyshev to show that each of these occur with high probability.

To start, because H is essentially D-regular, we have $|E| = (1 \pm \delta_1)Dn/r$, so

$$\mathbb{E}[|E'|] = p|E| = (1 \pm \delta_1)\epsilon n/r.$$

We also have

$$\mathrm{Var}(|E'|) = p(1-p)|E| \leq 2\epsilon n/r.$$

Because $\mathrm{Var}(|E'|) = o(\mathbb{E}[|E'|])$, Chebyshev will be able to show that E' is close to $\mathbb{E}[E']$ with high probability. More precisely, Chebyshev's inequality implies

$$\Pr[||E'| - \mathbb{E}[|E'|] \geq \delta_1\sqrt{2\epsilon n/r} \cdot \sqrt{2\epsilon n/r}] \leq \frac{r}{2\delta_1^2 \epsilon n} \leq .01,$$

with this last step holding for n sufficiently large. Thus with probability at least .99,

$$|E'| = \mathbb{E}[|E'|] \pm 2\delta_1 \epsilon n/r = (1 \pm 3\delta_1)\epsilon n/r.$$

This shows that (a) occurs with high probability. To deal with (b), let us first get a grasp on $\mathbb{E}[|V'|]$. For $x \in V$, let $\mathbb{1}_x = 1$ if $x \notin \bigcup_{e \in E'} e$ and $\mathbb{1}_x = 0$ otherwise. With this we see $|V'| = \sum_x \mathbb{1}_x$, so by linearity of expectation it suffices to bound each of $\mathbb{E}[\mathbb{1}_x]$.

We will say that a vertex x is *good* if $\deg(x) = (1 \pm \delta)D$ and that it is *bad* otherwise. If x is bad we will simply use the trivial estimates $0 \leq \mathbb{E}[\mathbb{1}_x] \leq 1$. If x is good we have

$$\mathbb{E}[\mathbb{1}_x] = (1-p)^{\deg(x)} = (1 - \epsilon/D)^{(1\pm\delta)D} = (1 \pm \delta_3)e^{-\epsilon}, \tag{22}$$

where this last step used that $1-p$ is within a constant factor of e^{-p} for p sufficiently small and that δ is chosen to be sufficiently small in terms of ϵ (e.g. we can make sure that it's smaller than ϵ^{-1}).

Having at most δn bad vertices by (i) together with (22) implies $\mathbb{E}[|V|'] = (1 \pm \delta_4)e^{-\epsilon}n$. To compute the variance, we observe that

$$\mathrm{Var}[|V'|] = \sum_x \mathrm{Var}[\mathbb{1}_x] + \sum_x \sum_{y \neq x} \mathbb{E}[\mathbb{1}_x \mathbb{1}_y] - \mathbb{E}[\mathbb{1}_x]\mathbb{E}[\mathbb{1}_y]. \tag{23}$$

Because each $\mathbb{1}_x$ is an indicator random variable, we have

$$\sum_x \mathrm{Var}[\mathbb{1}_x] \leq \sum_x \mathbb{E}[\mathbb{1}_x] = \mathbb{E}[|V'|].$$

For the mixed terms of (23), we have for any x, y that

$$\begin{aligned}\mathbb{E}[\mathbb{1}_x \mathbb{1}_y] - \mathbb{E}[\mathbb{1}_x]\mathbb{E}[\mathbb{1}_y] &= (1-p)^{\deg(x)+\deg(y)-\deg(x,y)} - (1-p)^{\deg(x)+\deg(y)} \\ &\le (1-p)^{-\deg(x,y)} - 1 \le (1-\epsilon/D)^{-\delta D} - 1 \le e^{\epsilon\delta} - 1 \le \delta_5,\end{aligned}$$

where our bound on $\deg(x, y)$ used (iii). In total we find

$$\mathrm{Var}[|V'|] \le \mathbb{E}[V'] + \delta_5 n^2 \le \delta_6(\mathbb{E}[V']|)^2,$$

where this last step used that $\mathbb{E}[|V'|] = \Theta_\epsilon(n)$. By Chebyshev we can guarantee with probability at least .99 that

$$|V'| = (1 \pm \delta_7)\mathbb{E}[|V'|] = (1 \pm \delta_8)ne^{-\epsilon}.$$

Proving that condition (c) holds with high probability is a little more complicated, so we'll omit the full details[1]. Let us instead give a heuristic argument as to why (c) holds in expectation. We first condition on the event $x \in V'$, which means that no edge containing x is in E'. An edge $e \ni x$ survives in $H[V']$ only if every edge f with $e \cap f \neq \emptyset$ has $f \notin E'$. Because H is roughly linear and D-regular, there are about rD such edges f, but D of these (namely those containing x) are automatically not in E' since we conditioned on $x \in V'$. The remaining $(r-1)D$ edges are each included independently and with probability ϵ/D, so the probability that none are included is $(1 - \epsilon/D)^{(r-1)D} \approx e^{-(r-1)\epsilon}$, and summing this over all of the roughly D edges containing x gives the result.

Once we have shown that each of (a),(b),(c) holds with probability at least .99, then the probability that all of them hold is at least .97, so in particular some choice of E' exists which satisfies these conditions. □

[1]The argument is similar in spirit to that of (b): you define $\mathbb{1}_e = 1$ if e survives in $H[V']$ and $\mathbb{1}_e = 0$ otherwise. Then $\deg'(x)$ is just the sum of some of these indicator random variables, so one has to bound terms of the form $\mathbb{E}[\mathbb{1}_e]$ and $\mathbb{E}[\mathbb{1}_e \mathbb{1}_f]$. If e, f are "typical" edges then the computation of $\mathbb{E}[\mathbb{1}_e]$ and $\mathbb{E}[\mathbb{1}_e \mathbb{1}_f]$ are straightforward to estimate, and there are few terms involving e which are not typical, giving the result.

21 Forbidden Submatchings

In the previous section we sketched a proof of Pippenger's Theorem 20.1, which roughly says that nearly-regular hypergraphs H with small codegrees have matchings which cover almost all of the vertices of H. Very recently there have been two significant strengthenings of this result due independently to Delcourt and Postle [47] and to Glock, Joos, Kim, Kühn, and Lichev [79], with these results very roughly saying that not only can we find large matchings in low-codegree hypergraphs, but moreover these matchings can be chosen while avoiding some set of "forbidden" submatchings.

In the present section we will discuss three theorems of increasing power due to Delcourt and Postle [47], the applications of which are typically referred to as the "forbidden submatching method." We will also make a few passing remarks about [79] as relevant, with a more detailed discussion of their work known as the "conflict-free hypergraph method" being be made in the following chapter.

We will make heavy usage of the following notation here and in the next chapter. We recall that a hypergraph H is *r-bounded* if every edge has size at most r, and we let $H^{(k)} \subseteq H$ be the subhypergraph of H consisting of all the edges of size exactly k. Given a hypergraph H, we say that a hypergraph $\mathcal{F}$ is a *conflict hypergraph* of H if $V(\mathcal{F}) = E(H)$ and if every hyperedge of $\mathcal{F}$ is a matching of size at least 2. In order to avoid confusion, we will typically refer to the edges of H as "edges" and the edges of $\mathcal{F}$ as "hyperedges". Given a conflict hypergraph $\mathcal{F}$ of H, we say that a matching M of H is *$\mathcal{F}$-avoiding* if no hyperedge of $\mathcal{F}$ (which is just a set of edges of H) is contained in M.

21.1 Pippenger's Theorem with Forbidden Submatchings

We begin with the following analog of Pippenger's Theorem, which gives almost perfect matchings that are additionally $\mathcal{F}$-avoidiing for some appropriate $\mathcal{F}$.

Theorem 21.1 ([47])**.** *For all integers $r, r' \geq 2$ and real $\beta \in (0,1)$, there exists $\alpha, D_\beta > 0$ such that the following holds for all $D \geq D_\beta$: let H be an r-bounded hypergraph such that*

(H1) $\Delta_1(H) \leq D$, and

(H2) $\Delta_2(H) \leq D^{1-\beta}$.

Further, let $\mathcal{F}$ be an r'-bounded conflict hypergraph of H with

(F1) $\Delta_1(\mathcal{F}^{(k)}) \leq \alpha D^{k-1} \log D$ for all $2 \leq k \leq r'$,

(F2) $\Delta_\ell(\mathcal{F}^{(k)}) \leq D^{k-\ell-\beta}$ for all $2 \leq \ell < k \leq r'$, and

(F3) Every hyperedge of $\mathcal{F}$ has size at least 3.

Then there exists an $\mathcal{F}$-avoiding matching of H of size at least $(1 - D^{-\alpha})D^{-1}|H|$.

Note that if we take $\mathcal{F} = \emptyset$ in this theorem and if H is r-uniform and close to D-regular, then this essentially recovers Pippenger's Theorem with the slightly stronger hypothesis of $\Delta_2(H) \leq D^{1-\beta}$ for the codegrees of H.

Theorem 21.1 as stated is a slight weakening of [47, Corollary 1.17], the full version of which replaces (F3) with a condition allowing for hyperedges of size 2 which are "well behaved". We will not need this strengthening for our main applications, so we defer a discussion about this full version of Theorem 21.1 to Section 21.3. We note that a nearly identical statement of (the strengthened version of) Theorem 21.1 also appears in [79], though their theorem requires the extra (and typically mild) assumption that H is not too sparse.

Just as we elected not to go through the proof of Pippenger's Theorem in full in the previous chapter, we will similarly not go through the significantly more complicated proof of Theorem 21.1 here, instead focusing on only on how to apply this theorem in practice. Just as we did in the previous chapter, we begin with a problem in design theory.

Recall that a hypergraph is called a partial (n, q, k)-Steiner system if it has n-vertices, is q-uniform, and if every k-set of vertices is contained in at most one edge, and we call this hypergraph an (n, q, k)-Steiner system if we further have that every k-set is contained in exactly one edge. For a partial (n, q, k)-Steiner system, we define a (v, e)-*configuration* to be a set of e edges which span at most v vertices.

It turns out that every (n, q, k)-Steiner system contains a $((q-k)i + k + 1, i)$-configuration for every fixed $i \geq 2$ [81, Proposition 7.1]. In the simplest case of $q = 3, k = 2$ where we are guaranteed $(i + 3, i)$-configurations, Erdős [56] asked if for all $g \geq 2$ and sufficiently large n, there exist $(n, 3, 2)$-Steiner systems with no $(i + 2, i)$-configurations for all $2 \leq i \leq g$, which we can think of as "high girth" Steiner systems. This problem was eventually solved in the positive in breakthrough work of Kwan, Sah, Sawhney, and Smikin [108].

Following this, Glock, Kühn, Lo, and Osthus [81] asked if a similar phenomenon held for all (n, q, k)-Steiner systems. That is, do there exist "high girth" Steiner systems with arbitrary parameters q, k? A not too difficult application of Theorem 21.1 gives the asymptotic version of this result, which was proven independently by Delcourt and Postle [47] and by [79].

Theorem 21.2 ([47, 79])**.** *For all $q > k \geq 2$ and $g \geq 2$, there exist partial (n, q, k)-Steiner systems $\mathcal{S}$ with*

$$|\mathcal{S}| \geq (1 - o(1)) \binom{q}{k}^{-1} \binom{n}{k}$$

and which contains no $((q-k)i + k, i)$-configuration for all $2 \leq i \leq g$.

Note that this in particular implies Rödl's Theorem 20.2 on asymptotically large designs.

Proof. As in the proof of Theorem 20.2, we start by defining H to be the $\binom{q}{k}$-uniform hypergraph whose vertex set is $\binom{[n]}{k}$ and which for every $Q \in \binom{[n]}{q}$, there exists a hyperedge $h_Q = \{K \in \binom{[n]}{k} : K \subseteq Q\}$. In this way, partial Steiner systems are exactly matchings in H. To incorporate Steiner systems with large girth, we (naively) define the conflict hypergraph $\mathcal{F}_0$ which consists of hyperedges F corresponding to $((q-k)i + k, i)$-configurations. More precisely, we define $\mathcal{F}_0$ to contain all matchings F of H with $2 \leq |F| \leq g$ such that $\left|\bigcup_{h_Q \in F} Q\right| \leq (q-k)|F| + k$.

With these definitions, we have that partial Steiner systems of large girth correspond exactly to $\mathcal{F}_0$-avoiding matchings in H. However, there is some unnecessary redundancy in how we defined $\mathcal{F}_0$. Namely, if we let $\mathcal{F} \subseteq \mathcal{F}_0$ consist of those F such that there exists no $F' \subseteq F$ of size at least 2 with $\left|\bigcup_{h_Q \in F'} Q\right| \le (q-k)|F'|+k$, then it is not difficult to check that we still have that partial Steiner systems of large girth correspond exactly to $\mathcal{F}$-avoiding matchings in H. Thus, while in principle we could solve the problem by working with either conflict hypergraph, we will only be able to verify the degree conditions in Theorem 21.1 for the smaller hypergraph $\mathcal{F}$. As we will see in later applications, this need to reduce to a slightly less straightforward conflict hypergraph with better degree conditions is a common component of implementing the forbidden submatching method in practice.

In total, we have now reduced our problem to finding an $\mathcal{F}$-avoiding matching in H that asymptotically covers all of its vertices, and for this we aim to apply Theorem 21.1. To this end, we fix some sufficiently small $\beta \in (0,1)$ (say $\beta = \frac{1}{2(q-k)}$, for example) and let $D_\beta, \alpha > 0$ be the numbers given by Theorem 21.1 for $r = \binom{q}{k}$, $r' = g$, and this value of β. As we will justify in a moment, we aim to apply Theorem 21.1 with $D = \binom{n-k}{q-k} = \Theta(n^{q-k})$, and for this we check the conditions of the theorem.

We first verify (H1), that $\Delta_1(H) \le D$. And indeed, every $K \in V(H)$ has $\deg_H(K) = \binom{n-k}{q-k} = D$.

We next verify (H2), that $\Delta_2(H) \le D^{1-\beta}$. And indeed, it is not difficult to check that distinct $K, K' \in v(H)$ satisfy $\deg_H(\{K, K'\}) \le \binom{n-k-1}{q-k-1} \le D^{1-\beta}$ provided n is sufficiently large in terms of β.

We next verify (F1), that $\Delta_1(\mathcal{F}^{(i)}) \le \alpha D^{i-1} \log D$ for all $2 \le i \le g$. Let h_Q be an edge of H. Observe that by definition of $\mathcal{F}$, any $F \in \mathcal{F}^{(i)}$ containing h_Q can be identified by first choosing the at most $(q-k)i + k - q$ elements in $[n] \setminus Q$ that are contained in $\bigcup_{h_{Q'} \in F} Q' \setminus Q$, and then choosing some $i-1$ subsets Q_j contained in this set of at most $(q-k)i + k \le 2qg$ elements of $[n]$. In total then the number of choices is at most $n^{(q-k)i+k-q} \cdot 2^{2qg} \le \alpha D^{i-1} \log D$ for n sufficiently large, and hence we conclude that the degree of any edge of H in $\mathcal{F}$ is at most this amount, verifying this condition.

We next verify (F2), that $\Delta_\ell(\mathcal{F}^{(i)}) \le D^{i-\ell-\beta}$ for all $2 \le \ell < i \le g$. Fix an arbitrary matching $M = \{h_{Q_1}, \ldots, h_{Q_\ell}\}$ of H for some $\ell \ge 2$. We first claim that if $\left|\bigcup_{j=1}^{\ell} Q_j\right| \le (q-k)\ell + k$, then $\deg_{\mathcal{F}^{(i)}}(M) = 0$ for all $i > \ell$. Indeed, in this case any $F \in \mathcal{F}$ of size i containing M would contain a subset $F' := M$ of size at least 2 satisfying $\left|\bigcup_{h_Q \in F'} Q\right| \le (q-k)|F'| + k$, contradicting the assumption $F \in \mathcal{F}$. Thus we may assume $\left|\bigcup_{j=1}^{\ell} Q_j\right| \ge (q-k)\ell + k + 1$. Similar to the (F1) case, this implies we can identify $F \in \mathcal{F}^{(i)}$ containing M by first choosing at most $(q-k)(i-\ell) - 1$ elements in $[n] \setminus \bigcup_{j=1}^{\ell} Q_j$, and then choosing some $i-1$ subsets Q_j of this set of at most $2qg$ vertices. In total this implies

$$\deg_{\mathcal{F}^{(i)}}(M) \le n^{(q-k)(i-\ell)-1} \cdot 2^{2qg} \le D^{i-\ell-\beta}$$

for n sufficiently large and $\beta < \frac{1}{q-k}$.

Finally, we verify (F3), that every hyperedge of $\mathcal{F}$ has size at least 3. Note that any $F = \{h_Q, h_{Q'}\} \in \mathcal{F}$ must have $|Q \cup Q'| \le 2q - k$ by definition of $\mathcal{F}$. This implies Q, Q' intersect in a set of size at least k, and hence $h_Q, h_{Q'}$ intersect in H, contradicting $F \in \mathcal{F}$ being a matching. We conclude that every hyperedge of $\mathcal{F}$ has size at least 3, verifying (F3).

By Theorem 21.1, there exists an $\mathcal{F}$-avoiding matching of H of size at least

$$(1 - D^{-\alpha})\frac{|H|}{D} = (1 - o(1))\frac{\binom{n}{q}}{\binom{n-k}{q-k}} = (1 - o(1))\binom{q}{k}^{-1}\binom{n}{k}.$$

This $\mathcal{F}$-avoiding matching corresponds to a partial Steiner system of high girth of the same size, proving the result. □

To get a better understanding of this proof and using the method, the reader may want to ask themselves the following questions:

- Where would the proof breakdown if we worked with $\mathcal{F}_0$ instead of $\mathcal{F}$?
- Where would the proof breakdown if we tried to find asymptotically large partial Steiner systems without $((q-k)i + k - 1, i)$-configurations (which we know by [81] always will exist in sufficiently large partial Steiner systems)? How large of a partial Steiner system can you find using this method without any of these configurations?

21.2 Bipartite Perfect Matchings

Theorem 21.1 turns out to be a consequence of an even stronger hypergraph matching result which guarantees not only a matching covering most of the vertices of H, but in fact a perfect matching provided our hypergraph is "bipartite". To this end, we say that a hypergraph $\mathcal{B}$ is *bipartite* if there exists an ordered partition (B_1, B_2) of the vertex set of $\mathcal{B}$ such that every edge of $\mathcal{B}$ intersects B_1 in exactly 1 vertex, and we refer to such a partition as a *bipartition*. In this setting we say that a matching M of $\mathcal{B}$ is a B_1*-perfect matching* if every vertex of B_1 is contained in exactly one edge of M.

Theorem 21.3 ([47] Theorem 1.16, Abridged)**.** *For all integers $r, r' \geq 2$ and real $\beta \in (0, 1)$, there exists $\alpha, D_\beta > 0$ such that the following holds for all $D \geq D_\beta$: Let $\mathcal{B}$ be a bipartite r-bounded hypergraph with bipartition (B_1, B_2) such that*

(B1) Every vertex in B_1 has degree at least $(1 + D^{-\alpha})D$ and every vertex in B_2 has degree at most D, and

(B2) $\Delta_2(\mathcal{B}) \leq D^{1-\beta}$.

Further, let $\mathcal{F}$ be an r'-bounded conflict hypergraph of $\mathcal{B}$ with

(F1) $\Delta_1(\mathcal{F}^{(k)}) \leq \alpha D^{k-1} \log D$ for all $2 \leq k \leq r'$,

(F2) $\Delta_\ell(\mathcal{F}^{(k)}) \leq D^{k-\ell-\beta}$ for all $2 \leq \ell < k \leq r'$, and

(F3) Every hyperedge of $\mathcal{F}$ has size at least 3.

Then there exist an $\mathcal{F}$-avoiding A-perfect matching of $\mathcal{B}$.

It is not immediately obvious that Theorem 21.3 strengthens Theorem 21.1. The key idea for this reduction comes from the following coloring result. Here we recall that the line graph $L(H)$ of a hypergraph H is the graph with vertex set $E(H)$ and where two edges of H are adjacent in $L(H)$ if and only if they intersect. We also define a proper t-coloring of a hypergraph to be a partition of its vertices into sets $V_1, \ldots, V_t$ so that no hyperedge is contained in any V_i set, and we define the chromatic number $\chi(H)$ to be the smallest t such that there exists a proper t-coloring.

Corollary 21.4. *For all integers $r, r' \geq 2$ and real $\beta \in (0,1)$, there exists $\alpha, D_\beta > 0$ such that the following holds for all $D \geq D_\beta$: let H be an r-bounded hypergraph such that*

(H1) $\Delta_1(H) \leq D$, and

(H2) $\Delta_2(H) \leq D^{1-\beta}$.

Further, let $\mathcal{F}$ be an r'-bounded conflict hypergraph of H with

(F1) $\Delta_1(\mathcal{F}^{(k)}) \leq \alpha D^{k-1} \log D$ for all $2 \leq k \leq r'$,

(F2) $\Delta_\ell(\mathcal{F}^{(k)}) \leq D^{k-\ell-\beta}$ for all $2 \leq \ell < k \leq r'$, and

(F3) Every hyperedge of $\mathcal{F}$ has size at least 3.

Then $\chi(L(H) \cup \mathcal{F}) \leq (1 + D^{-\alpha})D$.

It is not difficult to check that Theorem 21.1 follows from this by considering the matching which is the largest color-class in a proper $(1 + D^{-\alpha})D$-coloring of $L(H) \cup \mathcal{F}$, so it remains to prove this reduction.

Proof. Let C be a set of $(1 + D^{-\alpha})D$ colors. The main idea is that a coloring of the vertices $V(L(H)) = E(H)$ is just an assignment of each edge of $E(H)$ to exactly one value in C, and we will encode this assignment by a matching which covers each element of $E(H)$ exactly once (while making sure to add in some extra constraints to make sure this coloring is proper and avoids forbidden submatchings).

With this in mind, we define an auxilliary bipartite hypergraph $\mathcal{B}$ with bipartition (B_1, B_2) where $B_1 = E(H)$, $B_2 = \{(v,c) : v \in V(H),\ c \in C\}$, and with hyperedges of the form $h_{e,c} := \{e\} \cup \{(v,c) : v \in e\}$ for every $e \in H$ and $c \in C$. Observe that B_1-perfect matchings in $\mathcal{B}$ correspond exactly to $|C|$-proper colorings of $L(H)$, since each (v,c) vertex appearing in exactly one edge implies that no two edges e, e' which both contain v (and hence are adjacent in $L(H)$) are given the same color. Finally, we define an auxilliary forbidden hypergraph $\mathcal{F}'$ to consist of all hyperedges of the form $(F,c) := \{h_{e,c} : e \in F\}$ with $F \in \mathcal{F}$ and $c \in C$, and one can check that $\mathcal{F}'$-avoiding B_1-perfect matchings exactly correspond to proper $|C|$-colorings of $L(H) \cup \mathcal{F}$.

With the above in mind, it remains to check that an $\mathcal{F}$-avoiding B_1-perfect matching exists, for which we need to verify the conditions of Theorem 21.3.

We first verify (B1), that every vertex in B_1 has degree at least $(1 + D^{-\alpha})D$ and that every vertex in B_2 has degree at most D. And indeed, every vertex in B_1 has degree exactly $|C| = (1 + D^{-\alpha})D$, and every vertex $(v, c) \in B_2$ has degree at most $\deg_H(v) \le \Delta_1(H) \le D$, giving this condition.

We next verify (B2), that $\Delta_2(\mathcal{B}) \le D^{1-\beta}$. Indeed, any two vertices of B_1 have codegree 0, and any vertex in B_1 and vertex in B_2 have codegree at most 1, so the only case left to consider is when we have $(u, c), (v, c') \in B_2$. Their codegree will be 0 if $c \neq c'$, and if $c = c'$ then we have

$$\deg_{\mathcal{B}}(\{(u, c), (v, c)\}) = \deg_H(\{u, v\}) \le \Delta_2(H) \le D^{1-\beta},$$

giving this condition.

Finally, we observe that $\mathcal{F}'$ is isomorphic to $|C|$ disjoint copies of $\mathcal{F}$, and as such all the degree conditions for $\mathcal{F}$ carry over to $\mathcal{F}'$, verifying these last three conditions. We conclude that we can indeed apply Theorem 21.3, giving the desired $\mathcal{F}$-avoiding B_1-perfect matching.

□

The idea in the proof above about interpreting B_1-perfect matchings as colorings of the elements of B_1 can be used to solve other types of coloring problems. In particular, this observation (as well as further ideas in the study of conflict-free matchings) has been used recently to make tremendous progress around the generalized Ramsey problem of Erdős and Gyárfás which we introduced earlier in Theorem 3.5 around the local lemma. We highlight one particularly nice application of Theorem 21.3 to this area below, for which we recall some basic definitions.

A (p, q)-*coloring* of a graph G is an edge-coloring of G such that every p-clique of G receives at least q distinct colors, and we define the generalized Ramsey number $GR(n, p, q)$ to be the smallest number of colors needed in a (p, q)-coloring of K_n. In Theorem 3.5 we proved a result of Erdős and Gyárfás [58] showing the general bound

$$GR(n, p, q) \le p^{\frac{p^2}{\binom{p}{2} - q + 1}} n^{\frac{p-2}{\binom{p}{2} - q + 1}}.$$

While it is unknown how effective this bound is in general, it was shown in [58] that in the special case of $q = \binom{p}{2} - p + 3$, we have $GR(n, p, q) \ge \frac{n-1}{p-2}$, which matches the order of magnitude in the bound above of (very roughly) $GR(n, p, q) \le p^p n$. Because of this, the value

$$q_{lin} := \binom{p}{2} - p + 3$$

is known as the *linear threshold* for p, and a lot of the work around generalized Ramsey theory has been dedicated to determining what $GR(n, p, q_{lin})$ is asymptotically for various values of p.

Some of the initial progress towards understanding $GR(n, p, q_{lin})$ came about from the powerful but technical tools of the *differential equation method*, which led to very complicated proofs.

Later it was realized that these same results (and many more) could be achieved with much simpler proofs through the use of the matching methods of [47, 79]. Here we highlight one particularly nice application due to Bennett, Cushman, and Dudek [20] which improves the upper bound of roughly $GR(n, p, q_{lin}) \le p^p n$ from Erdős and Gyárfás's original paper to an asymptotic bound independent of p.

Theorem 21.5 ([20]). *For all $p \ge 3$, we have $GR(n, p, q_{lin}) \le (1 + o(1))n$.*

Unwinding the definitions here, this theorem calls for constructing some edge-coloring of K_n which avoids certain properties. Similar to how we proved Corollary 21.4, we will prove this result by constructing an auxiliary bipartite hypergraph $\mathcal{B}$ whose B_1-perfect matchings correspond to edge-colorings of K_n and a conflict hypergrpah $\mathcal{F}$ which sufficiently captures the properties we need to avoid.

Proof. When $p = 3$ we have $q_{lin} = 3$, so we are simply looking for the smallest number of colors needed to guaranteethat every triangle of K_n has 3 distinct colors, which is equivalent to asking for a proper edge-coloring of K_n. Vizing's theorem says this can be achieved using at most n colors, giving the result in this case. As such we assume $p \ge 4$ from now on.

Let C be a set of $(1 + o(1)n$ colors, with the exact size of C to be determined later. Exactly mimicing the proof of Corollary 21.4, we define a bipartite hypergraph $\mathcal{B}$ with $B_1 = E(K_n)$ and $B_2 = \{(v, c) : v \in K_n,\ c \in C\}$ with hyperedges $h_{uv,c} = \{uv\} \cup \{(u, c), (v, c)\}$ for all edges $uv \in E(K_n)$ and $c \in C$. Note that with this, matchings M in $\mathcal{B}$ correspond exactly to partial proper colorings χ_M of the edges of K_n, with B_1-perfect matchings exactly correspond to proper edge-colorings of all the edges of K_n. Here the most natural (though ultimately incorrect) conflict hypergraph $\mathcal{F}_0$ is defined by including all the hyperedges F whose corresponding partial coloring χ_F color the edges of some p-clique of K_n using at most $q_{lin} - 1 = \binom{p}{2} - p + 2$ distinct colors. As in our proof of Theorem 21.2, we will eventually need to replace $\mathcal{F}_0$ with a smaller conflict hypergraph $\mathcal{F}$, but for the moment we stick with this definition and see what goes wrong along the way.

Again, we emphasize that finding a (p, q)-coloring of K_n using at most $|C|$ colors is *exactly* the same as finding an $\mathcal{F}_0$-avoiding B_1-perfect matching in $\mathcal{B}$. As such, we should start by checking the conditions of Theorem 21.3 and see if we can conclude the result this way. To this end, we let $\beta \in (0, 1)$ be arbitrary and let $D_\beta, \alpha > 0$ be the values guaranteed from Theorem 21.3 with $r = 3$ and $r' = \binom{p}{2}$. We would like to show that $\mathcal{B}, \mathcal{F}_0$ satisfy the conditions of Theorem 21.3 for some appropriate value $D \ge D_\beta$, and (as will be justified shortly) we will try doing this with

$$D := n,$$

and we will further specify $C := n + n^{-\alpha}$ for future convenience. We begin by verifying the conditions for $\mathcal{B}$.

Claim 21.6. *With the parameters above, the hypergraph $\mathcal{B}$ satisfies conditions (B1) and (B2) of Theorem 21.3.*

Proof. We first check (B1), that every vertex in B_1 has degree at least $(1 + D^{-\alpha})D$ and every vertex in B_2 has degree at most D. For B_1, we note that the degree of every vertex $uv \in B_1$ is exactly

$$|C| = n + n^{-\alpha} = (1 + D^{-\alpha})D,$$

verifying this part of the condition. Moreover, every $(v, c) \in B_2$ is in exactly one edge of $\mathcal{B}$ for every $u \in V(K_n) \setminus \{v\}$, namely the edge $\{uv, (u, c), (v, c)\}$, so there degrees are all $n - 1 \leq D$, finishing the verification of this property. Moreover, this analysis for B_2 shows that for our given choice of hypergraphs, we could not possibly take D to be asymptotically smaller than n, which in turn shows with our analysis of B_1 that our set of colors C must have size at least roughly $n + n^{-\alpha}$, justifying our choices of parameters and explaining the ultimate conclusion of our theorem.

We next check (B2), that $\Delta_2(\mathcal{B}) \leq D^{1-\beta}$. It is easy to check that $\mathcal{B}$ is in fact linear, so this holds automatically. □

From here it is not difficult to check that $\mathcal{F}_0$ satisfies conditions (F1) and (F3) of **??**, but it fails miserably with (F2), as we verify in detail with the following claim. We note that this claim is entirely optional and only serves to help motivate our later choice of conflict-hypergraph $\mathcal{F}$.

Claim 21.7. *For $\ell = \binom{p}{2} - 1$, we have $\Delta_\ell(\mathcal{F}_0^{(\ell+1)}) \geq n$. In particular, we do not have $\Delta_\ell(\mathcal{F}_0^{(\ell+1)}) \leq D^{1-\beta}$.*

Proof. Because $\mathcal{F}_0$ is $\binom{p}{2}$-uniform, this statement just boils down to showing $\Delta_\ell(\mathcal{F}_0) \geq n$. To this end, consider some p-clique together with a proper coloring of $\binom{p}{2} - 1$ of its edges with the property that the coloring uses at most $q_{lin} - 2$ distinct colors (which exists, for example, by using Vizing's theorem to guarantee a proper coloring of these edges using at most $p \leq q_{lin} - 2$ colors for $p \geq 4$). Note that this partial coloring corresponds to some matching M in $\mathcal{B}$ of size ℓ. Because any of the $|C|$ choices for how to color the missing edge of this p-clique gives a p-clique using at most $q_{lin} - 1$ distinct colors, each of these $|C|$ choices corresponds to a forbidden hyperedge of $\mathcal{F}_0$ containing M. We conclude that $\Delta_\ell(\mathcal{F}_0) \geq \deg_{\mathcal{F}_0}(M) \geq |C| \geq n$. □

Similar to what we did in Theorem 21.2, our goal now is to construct a "smaller" conflict-hypergraph $\mathcal{F}$ such that $\mathcal{F}$-avoiding matchings are also $\mathcal{F}_0$-avoiding (so that these matchings still correspond to (p, q)-colorings). This will be achieved if we choose $\mathcal{F}$ such that it is an "undercover" of $\mathcal{F}_0$, i.e. if for all $F_0 \in \mathcal{F}_0$ there exists some $F \in \mathcal{F}$ with $F \subseteq F_0$.

The idea for how to choose such an undercover $\mathcal{F}$ comes in part from the proof of the claim. There we saw that a major obstruction to verifying (F2) for $\mathcal{F}_0$ was the existence of partial colorings of most of the edges of a p-clique which use very few distinct colors, and in particular so few that no completion of our coloring could possibly work. The core idea now is that we will add to $\mathcal{F}_0$ all of the "minimal" partial colorings with the property that no completion can possibly work and then remove from $\mathcal{F}_0$ any colorings that contain this, thereby eliminating from $\mathcal{F}_0$ all of its "very bad" edges while maintaining the desired undercover property.

To this end, given a matching M, we let $v(M)$ denote the number of vertices of K_n incident to a colored edge of the associated coloring χ_M, and we let $c(M)$ denote the total number of colors used by χ_M. We then define $\mathcal{F}$ to consist of all matchings F of $\mathcal{B}$ with $3 \leq v(F) \leq p$ such that

$$\rho(F) := |F| - c(F) - v(F) + 2 \geq 0,$$

and further such that there exists no submatching $F' \subseteq F$ with $v(F') \geq 3$ satisfying $\rho(F') \geq 0$. It should not be obvious at this point why we have defined ρ exactly as we have. The very rough intuition here is that the total number of matchings M with a given value of $c(M)$ and $v(M)$ is roughly $n^{c(M)+v(M)}$ (since we have about n ways to choose each color and each vertex used). As such, if we want $\mathcal{F}$ to not have too many elements (and hence not too large of a degree like we had before), then we need to omit from $\mathcal{F}$ any F with $c(F)+v(F)$ large. Our choice of ρ achieves this since it omits from $\mathcal{F}$ any F with $\rho(F) < 0$, i.e. with $c(F)+v(F) > |F|+2$. Our exact choice of $+2$ as a normalizing factor is done so that $\rho(F) \geq 0$ whenever $v(F) = p$ and $c(F) = q_{lin} - 1$, which is implicitly the key fact we use in the following important observation.

Claim 21.8. *If M is an $\mathcal{F}$-avoiding B_1-perfect matching, then χ_M is a (p, q_{lin})-coloring.*

Proof. Assume for contradiction that there existed some p-clique whose edges contained at most $q_{lin} - 1$ distinct colors under χ_M. Letting $M' \subseteq M$ denote the matching restricted to the edges of this p-clique, we see that

$$\rho(M') = \binom{p}{2} - c(M') - p + 2 \geq \binom{p}{2} - q_{lin} + 1 - p + 2 = 0.$$

Letting $F \subseteq M'$ be any minimal subset with the properties $v(F) \geq 3$ and $\rho(F) \geq 0$ (which exists by the computation above), we have by definition that $F \subseteq M$ is a hyperedge of $\mathcal{F}$, contradicting our assumption of M being $\mathcal{F}$-avoiding. $\square$

With this and our first claim, we see by Theorem 21.3 that it suffices to verify that $\mathcal{F}$ satisfies conditions (F1), (F2), and (F3).

We first verify (F1), that $\Delta_1(\mathcal{F}^{(k)}) \leq \alpha D^{k-1} \log D$ for all $2 \leq k \leq \binom{p}{2}$. Fix some 3-edge h in $\mathcal{B}$ (which is a vertex in $\mathcal{F}$) and an integer k, and for each integer $1 \leq t \leq p$, let $\mathcal{F}_t^{(k)}(h)$ denote the set of hyperedges $F \in \mathcal{F}^{(k)}$ containing h with $v(F) = 2 + t$.

Claim 21.9. *We have $|\mathcal{F}_t^{(k)}(h)| \leq k^{3k^2} n^{k-1}$ for all $1 \leq t \leq p$.*

Proof. Let $h = \{uv, (u,c), (v,c)\}$. We first observe that because every $F \in \mathcal{F}_t^{(k)}(h)$ in particular has $\rho(F) \geq 0$, we must have $c(F) \leq k - t$ by definition of ρ, and we note that one of these at most $k - t$ colors must be c since $h \in F$.

With this in mind, we can identify each $F \in \mathcal{F}_t^{(k)}(h)$ as follows: first choose the t additional vertices other than u, v which are incident to colored edges of χ_F in at most n^t ways, then choose the at most $k - t - 1$ colors other than c used by χ_F in at most $|C|^{k-t-1}$ ways, then finally choose $k-1$ new edges to add to M which are incident to our chosen set of $2+t$ vertices and color them using the at most k colors that we have allowed (the number of ways for which can very roughly be estimated to be at most $k^{2k(k-1)} \leq k^{2k^2}$). Putting all this together and using $|C| \leq 2n \leq kn$ for n sufficiently large gives the result. $\square$

Since $\deg_{\mathcal{F}^{(k)}}(h) = \sum_t |\mathcal{F}_t^{(k)}(h)| \leq pk^{3k^2} n^{k-1} \leq \alpha D^{k-1} \log D$ for n sufficiently large, we conclude that (F1) holds.

We next verify (F2), that $\Delta_\ell(\mathcal{F}^{(k)}) \leq D^{k-\ell-\beta}$ for all $2 \leq \ell < k \leq \binom{p}{2}$. Fix some matching M of size $\ell \geq 2$ (noting that this implies $v(M) \geq 3$) and some integer $k > \ell$. Observe that if

$\rho(M) \geq 0$ then $\deg_{\mathcal{F}^{(k)}}(M) = 0$, as any F of size k containing M would in particular contain the submatching $F' = M$ with $v(F') \geq 3$ and $\rho(F') \geq 0$, meaning $F \notin \mathcal{F}$ by definition. Thus we may assume $\rho(M) < 0$. Similar to before, for all $0 \leq t \leq p$ we let $\mathcal{F}_t^{(k)}(M)$ denote the set of $F \in \mathcal{F}$ of size k containing M with $v(F) = v(M) + t$.

Claim 21.10. *We have $|\mathcal{F}_t^{(k)}(M)| \leq k^{3k^2} n^{k-\ell-1}$ for all $0 \leq t \leq p$.*

Proof. Because $\mathcal{F}_t^{(k)}(M) \subseteq \mathcal{F}$, we must have $\rho(F) \geq 0$ for all $F \in \mathcal{F}_t^{(k)}(M)$, which by our assumption $\rho(M) < 0$ above implies

$$0 < \rho(F) - \rho(M) = [k - c(F) - v(M) - t + 2] - [\ell - c(M) - v(M) + 2] = k - \ell - t - c(F) + c(M),$$

or equivalently $c(F) \leq c(M) + k - \ell - t - 1$ for all $F \in \mathcal{F}_t^{(k)}(M)$. As we will see in a moment, this is the other key fact we need about our exact definition of ρ.

We now identify each $F \in \mathcal{F}_t^{(k)}(M)$ as follows: first choose the set of colors used by χ_F that do no appear in χ_M (which can be done in at most $|C|^{k-\ell-t-1}$ ways by our calculation above), then choose the t vertices that are incident to colored edges of χ_F that are not incident to χ_M in at most n^t ways, then finally choose $k - \ell$ new edges to add to M which are incident to our chosen set of $v(M) + t$ vertices and color them using the at most k colors that we have allowed (which can very roughly be estimated to be at most $k^{2k(k-\ell)} \leq k^{2k^2}$). Putting all this together and using $|C| \leq 2n \leq kn$ for n sufficiently large gives the result. $\square$

Since $\deg_{\mathcal{F}^{(k)}}(M) = \sum_t |\mathcal{F}_t^{(k)}(M)| \leq (p+1)k^{3k^2} n^{k-\ell-1} \leq D^{k-\ell-\beta}$ for n sufficiently large, we conclude that (F2) holds.

Finally we verify (F3), that $\mathcal{F}$ contains no edges of size 3. This is immediate from how we defined $\mathcal{F}$, so we conclude that Theorem 21.3 applies, giving the desired result. $\square$

Again, the interested reader is encouraged to think about where this proof would have broken down if we tried proving a linear upper bounding on $GR(n, p, q)$ for some $q > q_{lin}$.

21.3 Matchings of Size 2

While Theorem 21.3 is often enough for applications of the forbidden submatching method, sometimes we need to consider conflict hypergraphs $\mathcal{F}$ which have matchings of size 2. To deal with this, we need to introduce some particular notions of degrees involving edges of size 2.

To this end, if $\mathcal{F}$ is a conflict hypergrpah of a hypergraph H, then for $v \in V(H)$ and $e \in E(H)$ with $v \notin e$, we define their *mixed-codegree* to be the number of hyperedges $F \in \mathcal{F}$ of size 2 which contains e and another edge e' which contains v. For two distinct vertices $e, e' \in V(\mathcal{F}) = E(H)$, we define their *common 2-degree* to be the number of vertices e'' such that $\{e, e''\}, \{e', e''\}$ are both hyperedges in $\mathcal{F}$.

The following result is exactly the same as Theorem 21.3 except for the modified (F3) condition and the slightly stronger conclusion about multiple disjoint matchings.

Theorem 21.11 ([47] Theorem 1.16)**.** *For all integers $r, r' \geq 2$ and real $\beta \in (0,1)$, there exists $\alpha, D_\beta > 0$ such that the following holds for all $D \geq D_\beta$: let $\mathcal{B}$ be a bipartite r-bounded hypergraph with bipartition (B_1, B_2) such that*

(B1) Every vertex in B_1 has degree at least $(1 + D^{-\alpha})D$ and every vertex in B_2 has degree at most D, and

(B2) $\Delta_2(\mathcal{B}) \leq D^{1-\beta}$.

Further, let $\mathcal{F}$ be an r'-bounded conflict hypergraph of $\mathcal{B}$ with

(F1) $\Delta_1^{(k)}(\mathcal{F}) \leq \alpha D^{k-1} \log D$ for all $2 \leq k \leq r'$,

(F2) $\Delta_\ell^{(k)}(\mathcal{F}) \leq D^{k-\ell-\beta}$ for all $2 \leq \ell < k \leq r'$,

(F3') The maximum mixed-codegree and maximum common 2-degree of $\mathcal{F}$ is at most $D^{1-\beta}$.

Then there exist at least D disjoint $\mathcal{F}$-avoiding B_1-perfect matchings of $\mathcal{B}$.

Again, we note that essentially the same theorem appears in [79] with the additional mild hypothesis that H is not too sparse. The conclusion of Theorem 21.11 that $\mathcal{B}$ contains many disjoint perfect matchings is rarely used in practice, but it is certainly a neat fact that you can guarantee all of these exist.

One place where the full power of Theorem 21.11 is needed over the simpler Theorem 21.3 is in proving the following improvement to the Erdős-Gyárfás bound Theorem 3.5 for generalized Ramsey numbers by a logarithmic factor for a wide range of p, q. We note that, at least of this time of writing, this is the only known order of magnitude improvements for the original bound of Erdős and Gyárfás.

Theorem 21.12 ([21, 22])**.** *If p, q are integers such that $p-2$ is not divisible by $\binom{p}{2} - q + 1$, then*

$$GR(n,p,q) = O\left(\left(\frac{n^{p-2}}{\log n}\right)^{\frac{1}{\binom{p}{2}-q+1}}\right).$$

This result was originally proven in the range $q \leq p^2/4$ by Bennett, Dudek, and English [22] using the differntial equations method, after which the proof was significantly simplified and extended to the full range of q by Bennett, Delcourt, Li, and Postle [21] using the forbidden submatching method.

Sketch of Proof. One immediate observation is that for most values of q in the theorem, the bound we are shooting for here is $o(n)$. As such, there is no hope in us proving the existence of a *proper* coloring which uses at most this many colors. In particular, there is no hope in us using a bipartite hypergraph like we did in the proof of Theorem 21.5 which had hyperedges of the form $\{uv, (u,c), (v,c)\}$ since in matchings in this hypergraph correspond to proper colorings of K_n.

Getting around this first obstacle is not too difficult: we instead consider the bipartite (hyper)graph $\mathcal{B}$ with $B_1 = E(K_n)$, $B_2 = \{(e,c) : e \in E(K_n),\ c \in C\}$, and all edges of the form $\{e,(e,c)\}$ (so $\mathcal{B}$ is just a graph which is the union of stars). From here we can mimic the exact proof strategy we had before with Theorem 21.5 where we defined our conflict hypergraph $\mathcal{F}$ to consist of matchings of $\mathcal{B}$ which correspond to (minimal) colorings of K_n which have few vertices and few distinct colors. More precisely, we let $\mathcal{F}$ consist of all minimal matchings M with the property that

$$\rho(M) := |M| - c(M) - \frac{\binom{p}{2} - q + 1}{p-2}(v(M) - 2) \geq 0,$$

noting that in the special case of $q = q_{lin} = \binom{p}{2} - p + 3$ this exactly recovers our previous definition of ρ, and just like in this case we have $\rho(M) = 0$ whenever M corresponds to a coloring of all the edges of a p-clique using $q-1$ colors.

At this point, checking the conditions of our forbidden matching theorem largely proceed as before, with the critical exception being that $\mathcal{F}$ will now contain many hyperedge sof size 2, essentially because $\mathcal{B}$ no longer corresponds to proper colorings. Thus one has to additionally check the bounds of (F3') in our analysis, and this is not too difficult to do since one can check that we in fact have $\Delta_1(\mathcal{F}^{(2)}) \leq D^{1-\beta}$, implying (F3'). $\square$

22 TODOConflict-Free Matchings

Part VII

Linear Algebra Methods

Very roughly speaking, the *linear algebra method* in combinatorics works as follows:

1 Associate a "linear algebraic object" M to your problem (e.g. a matrix, list of vectors).

2. Determine algebraic information about M (e.g. its rank, eigenvalues, eigenvectors),

3. Use this algebraic information to say something about your original problem.

The linear algebra method applies to a broad range of problems. We only scratch the surface here, and we refer the reader to books by Babai and Frankl [10] and by Matoušek [113] for a more thorough treatment of this versatile method.

23 Modular Intersections

We begin with a very classical application of the linear algebra method: Oddtown.

Consider the following (somewhat whimsical) setup. The city of Oddtown has a variety of clubs, each of which follows the following odd set of rules: each club must have an odd number of people, and every two distinct clubs must have an even number of people in common.

The main question now becomes: if Oddtown has n people, what's the maximum number of clubs it can have? Equivalently, if $\mathcal{F} = \{F_1, F_2, \ldots, F_m\} \subseteq 2^{[n]}$ is a set system such that $|F_i|$ is odd for all i and such that $|F_i \cap F_j|$ is even for all $i \neq j$, then what is the maximum size of $\mathcal{F}$?

A very simple construction is to take $F_i = \{i\}$ for all i, which trivially satisfies the stated conditions. However, it's far from the only construction. For example, if n is even one can also take each F_i to be either $\{i\}$ or $[n] \setminus \{i\}$, and there are many, many more constructions achieving a bound of n (in fact, there's close to 2^{n^2} non-isomorphic constructions due to Szegedy [10, Exercise 1.1.14]).

Given all of these constructions, it seems plausible that (1) the true answer is indeed n, and (2) proving this might be difficult (since we have to come up with an argument that somehow deals with all of these constructions in a unified way). Fortunately, the linear algebra method manages to give a unified approach for all of these constructions in an extremely elegant way. More generally, we note that it's a good rule of thumb is that if there are many distinct looking constructions, then the linear algebra method might come in handy.

Theorem 23.1 (Oddtown)**.** *Let $\mathcal{F} \subseteq 2^{[n]}$ be a set system such that $|F|$ is odd for all $F \in \mathcal{F}$ and such that $|F \cap F'|$ is even for all $F \neq F' \in \mathcal{F}$. Then $|\mathcal{F}| \leq n$.*

Proof. Given a set $F \subseteq [n]$, define its characteristic vector $\chi_F \in \mathbb{F}_2^n$ by having $(\chi_F)_i = 1$ if $i \in F$ and $(\chi_F)_i = 0$ otherwise. Note crucially that for any F, F', the dot product satisfies

$$\langle \chi_F, \chi_{F'} \rangle = |F \cap F'| \mod 2.$$

We claim that $\{\chi_F : F \in \mathcal{F}\}$ is a set of linearly independent vectors. Indeed, say we had

$$\sum_{F \in \mathcal{F}} \lambda_F \chi_F = 0.$$

Take any $F' \in \mathcal{F}$ and apply the dot product on both sides to get

$$\sum_{F \in \mathcal{F}} \lambda_F \langle \chi_F, \chi_{F'} \rangle = 0.$$

By the observation above and the hypothesis of the theorem, we see $\langle \chi_F, \chi_{F'} \rangle = 0$ if $F \neq F'$ and that $\langle \chi_{F'}, \chi_{F'} \rangle = 1$. Thus the above says $\lambda_{F'} = 0$, and as $F' \in \mathcal{F}$ was arbitrary, we conclude that these vectors are indeed linearly independent.

Since we have $|\mathcal{F}|$ linearly independent vectors in $\mathbb{F}_2^n$, we must have $|\mathcal{F}| \leq n$, giving the result. $\square$

Since that was so clean, let's give another (essentially equivalent) proof of this result using slightly different language, which in some situations might be simpler to think about.

Proof. Write $\mathcal{F} = \{F_1, \ldots, F_m\}$ and let M be the $m \times n$ matrix over $\mathbb{F}_2$ which has $M_{i,j} = 1$ if $j \in F_i$ and $M_{i,j} = 0$ otherwise. Let $L = MM^T$. We claim that L is the $m \times m$ identity matrix. Indeed, one can verify that $L_{i,j} \equiv |F_i \cap F_j| \mod 2$, so the hypothesis of the theorem gives this claim.

Using the general fact

$$\text{rank}(AB) \le \text{rank}(A) \le \#\text{columns of } A,$$

which is valid for any two matrices A, B for which AB makes sense; we see that we can apply this with $A = B^T = M$ to conclude that

$$m = \text{rank}(I_m) \le \text{rank}(M) \le n,$$

proving the result. ☐

While the proof of Oddtown is extremely nice, one might complain that the problem itself is rather ad hoc. Here we give a far reaching generalization of the Oddtown problem which, in addition to being nice on its own, has a number of important applications.

To this end, we adopt the following (non-standard) notation. For an integer p and a set $L \subseteq \{0, 1, \ldots, p-1\}$, we say that a family $\mathcal{F}$ is (p, L)-*modular intersecting* if $|F| \mod p \notin L$ for all $F \in \mathcal{F}$ and if $|F \cap F'| \mod p \in L$ for all $F \neq F' \in \mathcal{F}$.

For example, a family is $(2, \{0\})$-modular intersecting if and only if it follows the rules of Oddtown. By following a somewhat similar strategy as in Oddtown, we can prove the following.

Proposition 23.2. *For any prime p and $L \subseteq \{0, 1, \ldots, p-1\}$, if $\mathcal{F} \subseteq 2^{[n]}$ is (p, L)-modular intersecting then*

$$|\mathcal{F}| \le \sum_{i=0}^{|L|} \binom{n+i-1}{i}.$$

Proof. Similar to before, we define the characteristic vector $\chi_F \in \mathbb{F}_p^n$ by having $(\chi_F)_i = 1$ if $i \in F$ and $(\chi_F)_i = 0$ otherwise, and we again observe that $\langle \chi_F, \chi_{F'} \rangle = |F \cap F'| \mod p$. Unfortunately we can't conclude that these vectors are linearly independent like we could in the Oddtown case, but we can do this if we generalize our vectors somewhat.

To this end, for each $F \in \mathcal{F}$, define the polynomial $p_F : \mathbb{F}_p^n \to \mathbb{F}_p$ by

$$p_F(\vec{x}) = \prod_{\ell \in L} (\langle \chi_F, \vec{x} \rangle - \ell).$$

Note that by hypothesis, we have $p_F(\chi_{F'}) = 0$ if $F \neq F' \in \mathcal{F}$ (since $\langle \chi_F, \chi_{F'} \rangle \equiv |F \cap F'| \equiv \ell \mod p$ for some $\ell \in L$) and that $p_F(\chi_F) \neq 0$ (since $|F| \not\equiv \ell \mod p$ for any $\ell \in L$).

We claim that the polynomials $\{p_F : F \in \mathcal{F}\}$ are linearly independent. Indeed, say we had

$$\sum_{F \in \mathcal{F}} \lambda_F p_F = 0.$$

By plugging in $\chi_{F'}$ for some $F' \in \mathcal{F}$ into both sides of this equality, the observation above implies $c_{F'}\lambda_{F'} = 0$ for some $c_{F'} \neq 0$, and hence $\lambda_{F'} = 0$, proving the claim.

Observe that each polynomial p_F has degree at most $|L|$. It is a basic exercise to show that the dimension of the space of polynomials in n variables with degree at most $|L|$ is equal to $\sum_{i=0}^{|L|} \binom{n+i-1}{i}$. Since we constructed a set of $|\mathcal{F}|$ linearly independent polynomials that lie in a space of dimension $\sum_{i=0}^{|L|} \binom{n+i-1}{i}$, we must have $|\mathcal{F}| \le \sum_{i=0}^{|L|} \binom{n+i-1}{i}$, proving the result. **Maybe add notation like $V_{n,\le d}$ for this vector space.** ☐

The bound of Proposition 23.2 is relatively close to tight. Indeed, if $L = \{0, 1, \dots, s-1\}$, then $\mathcal{F} = \binom{[n]}{s}$ is (p, L)-modular intersecting and has size $\binom{n}{s} \approx \sum_{i=0}^{s} \binom{n+i-1}{i}$. However, one can do better by refining our argument (though we emphasize that the present weaker bound of Proposition 23.2 is all one typically needs for most applications).

Towards refining this proof/bound, one might first turn to the simplest case of $p = 2,\ L = \{0\}$. As this is just Oddtown, we know the answer here should be n, but the bound we get is $n+1$. By carefully analyzing the proof, one realizes that the one place our argument isn't sharp is when we argue about the dimension of the space spanned by the p_F polynomials. Indeed, we naively said that they lived in the space of degree at most 1 polynomials, but in fact they are all in the span of the monomials $\{x_i\}$, which gives the optimal bound.

We are thus left with the problem of trying to find a smaller subspace for us to work in while achieving the same conclusion, and there are a couple of semi-standard ways of doing this. One way is to try and find polynomials which are more "efficient", i.e. of lower degree, for which our conclusion still holds. And indeed, if you look back through the proof of Proposition 23.2, one sees that although we defined our polynomials to have codomain $\mathbb{F}_p^n$, we only fed them 0-1 vectors to show linear independence. As such, it would suffice to look at polynomials $\bar{p}_F$ which agree with the p_F polynomials on 0-1 vectors, and this can be done simply by replacing each x_i^α in the expansion of p_F by x_i (i.e. by replacing p_F by a multilinear function). For example, if $p_F(\vec{x}) = 3x_1x_2^2 + 2x_1^3x_3^2$ then we would want to look at $\bar{p}_F(\vec{x}) = 3x_1x_2 + 2x_1x_3$. By using this we can prove the following.

Theorem 23.3. *For any prime p and $L \subseteq \{0, 1, \dots, p-1\}$, if $\mathcal{F} \subseteq 2^{[n]}$ is (p, L)-modular intersecting then*

$$|\mathcal{F}| \le \sum_{i=0}^{|L|} \binom{n}{i}.$$

We note that this theorem was originally proven by Deza, Frankl, and Singhi [50], with the following simpler proof due to Alon, Babai, and Suzuki [2].

Proof. Define p_F as before, and define the multilinear polynomial $\bar{p}_F$ by replacing each x_i^α in the expansion of p_F by x_i. Note that $p_F(\chi_{F'}) = \bar{p}_F(\chi_{F'})$ for all F', so by following the same proof as before we conclude that the $\bar{p}_F$ polynomials are linearly independent. Moreover, these linearly independent polynomials all lie in the space of multilinear polynomials in n variables with degree at most $|L|$, which is a space of dimension $\sum_{i=0}^{|L|} \binom{n}{i}$. We thus must have $|\mathcal{F}| \le \sum_{i=0}^{|L|} \binom{n+i-1}{i}$, proving the result. ☐

The reader might still complain that, although we have improved our bound, it still fails to be tight for Oddtown. As far as we are aware, it is still an open problem as to whether the present bound can be replaced by $|\mathcal{F}| \le \binom{n}{|L|}$ for all L, which would be tight whenever $L = \{0, 1, \ldots, s-1\}$. However, such a general result is known to hold when $\mathcal{F}$ is uniform.

Theorem 23.4 (Frankl-Wilson [67]). *For any prime p and $L \subseteq \{0, 1, \ldots, p-1\}$, if $\mathcal{F} \subseteq 2^{[n]}$ is (p, L)-modular intersecting and k-uniform, then*

$$|\mathcal{F}| \le \binom{n}{|L|}.$$

We omit their proof, which utilizes the powerful tool of incidence matrices, and instead refer the reader to [10, Chapter 7] for more on this. Instead, we will prove a weaker version of this result which holds in the non-modular setting.

For this, given a set of integers L, we say that a set system $\mathcal{F}$ is *L-intersecting* if $|F \cap F'| \in L$ for all $F \ne F' \in \mathcal{F}$.

Theorem 23.5 (Ray-Chaudhuri-Wilson [126]). *If L is a set of integers and $\mathcal{F} \subseteq 2^{[n]}$ is L-intersecting and k-uniform, then*

$$|\mathcal{F}| \le \binom{n}{|L|}.$$

Proof. Note that we may assume $L \subseteq \{0, 1, \ldots, k-1\}$ since these are the only intersections that can actually occur, and in particular we may assume $|L| \le k$.

Similar to before, we define polynomials $p_F : \mathbb{R}^n \to \mathbb{R}$ by $p_F(\vec{x}) = \prod_{\ell \in L}(\langle \chi_F, \vec{x}\rangle - \ell)$ and we define $\bar{p}_F$ to be their multilinearizations. As before, we find that these $\bar{p}_F$ polynomials are linearly independent, giving a bound of $\sum_{i=0}^{|L|} \binom{n}{i}$ since these polynomials lie in the vector space V of multilinear polynomials of degree at most $|L|$.

Observe that the bound above fails to be sharp if and only if these $\bar{p}_F$ do not span all of V. In particular, to prove the result, it suffices to find a set of $\sum_{i=0}^{|L|-1} \binom{n}{i}$ polynomials of V which are linearly independent with the $\bar{p}_F$ vectors.

Say we write these supposed polynomials as q_I for $I \subseteq [n]$ with $|I| \le |L| - 1$, what would be a good choice for these? Well, to mimic the previous proof we would like to have $q_I(\chi_F) = 0$ for all $F \in \mathcal{F}$. Since all we know about the individual elements of $\mathcal{F}$ is that they each of size k, perhaps a reasonable condition is to demand

$$q_I(\vec{x}) = q_I'(\vec{x}) \left(\sum_{j=1}^{n} x_j - k\right),$$

where q_I' is some appropriate polynomial of degree at most $|L| - 1$. Indeed, if we do this and if we have

$$\sum_F \lambda_F \bar{p}_F + \sum_I \lambda_I q_I = 0,$$

then by plugging in $\chi_{F'}$ into both sides of this equality we see that $\lambda_F = 0$ for all $F \in \mathcal{F}$. As such, all we have to do is choose the q_I' polynomials such that the q_I vectors themselves are linearly independent. One possibility is to set

$$q_I(\vec{x}) = \prod_{i\in I} x_i \cdot \left(\sum_{j=1}^{n} x_j - k\right).$$

Note that with this, for any set $J \subseteq [n]$ with $|J| \leq |L| - 1 < k$, we have $q_I(\chi_J) = 0$ if $I \not\subseteq J$, and we have $q_I(\chi_J) \neq 0$ if $I \subseteq J$ (here we implicitly use $|J| < k$ to ensure $\sum_{j=1}^{n}(\chi_J)_j - k \neq 0$). With this in mind, say we had

$$\sum_I \lambda_I q_I = 0,$$

and let J be a smallest set such that $\lambda_J \neq 0$. By plugging χ_J into both sides of this expression, the observation above implies that we get $c_J \lambda_J = 0$ for some $c_J \neq 0$, and hence $\lambda_J = 0$, a contradiction. We conclude that the q_I polynomials are linearly independent, and hence all of the polynomials $\{\bar{p}_F\}_F \cup \{q_I\}_I$ are linearly independent vectors. Unfortunately the q_I polynomials aren't in V, i.e. aren't multilinear, but this can be easily remedied by considering $\{\bar{p}_F\}_F \cup \{\bar{q}_I\}_I$, and again we conclude that these vectors (in V) are linearly independent, giving the desired bound on $|\mathcal{F}|$. $\square$

We note that the exact proof as written almost goes through if instead of demanding $|F| = k$ for $F \in \mathcal{F}$, we only demand $|F| \equiv k \mod p$. A careful reading shows that this will go through if we assume $k \notin [|L| - 1] \mod p$ (which holds if e.g. $|L| \leq k$), but without this assumption we can't assume $q_I(\chi_J) \neq 0$ whenever $I \subseteq J$. Nevertheless, it does turn out that this bound does continue to hold under the weaker hypothesis $|F| \equiv k \mod p$, but of the proof of this ends up being somewhat more involved and we refer the interested reader to [10, Theorem 5.37].

24 Pseudo-Adjacency Matrices

Given a graph G, we define its adjacency matrix A to be the matrix whose rows and columns are indexed by $V(G)$ and where $A_{i,j} = 1$ if $i \sim j$ and $A_{i,j} = 0$ otherwise. While A is a natural way to encode a graph G, it is not at all obvious that its spectral properties should tell you anything about G, but this remarkably does turn out to be the case! For example, we'll see below that $\sigma(A)$ can give effective bounds on its maximum degree, as well as its independence number.

Many of these results which hold for A (and which perhaps are most naturally proven by thinking about A) can be tweaked to hold for a wider class of "pseudo-adjacency matrices", which exhibit behavior similar to A but which are cooked up to deal with the specific problem at hand. We exhibit this phenomenon in the next two subsection. In each subsection, we begin by highlighting important lemmas relating the spectrum of A to G, after which we generalize these lemmas apply to a slightly broader class of pseduo-adjacency matrices M, and we then carefully choose such an M to solve a cool problem.

Before moving on, we note that much more broadly, the area of *spectral graph theory* concerns studying matrices M associated to graphs G and how the spectral properties of M relate to combinatorial properties of G. We will not have time to dive very deeply into this fascinating area, and we refer the interested reader to **the appendix** for more on this.

24.1 Huang's Theorem

In this subsection we give a simple proof of Huang's, which solved what used to be a 30 year old problem known as the sensitivity conjecture. Our proof relies on two basic results from linear algebra, both of which are used heavily throughout spectral graph theory. First we have the Rayleigh quotient, which gives an analytic way to compute the eigenvalues of a symmetric matrix. Here $\lambda_1(M)$ denotes the largest eigenvalue of M.

Lemma 24.1. *Let M be a real symmetric matrix. Then*

$$\lambda_1(M) = \max_{x \neq 0} \frac{x^* M x}{x^* x},$$

and any x achieving equality is an eigenvector corresponding to $\lambda_1(M)$.

As a small application **Actually do we even use the Raleigh quotient here?**, we use the Rayleigh quotient to establish a connection between eigenvalues and combinatorial information of graphs.

Lemma 24.2. *Let G be a graph and A its adjacency matrix. Then*

$$\lambda_1(A) \leq \Delta,$$

where Δ is the maximum degree of G.

Proof. let x be an eigenvector of A corresponding to $\lambda_1(A)$ and let $v \in V(G)$ be such that $|x_v|$ is maximized. Then by our definitions, we have

$$|\lambda_1(A)x_v| = |(Ax)_v| = |\sum_u A_{v,u}x_u| \leq \sum_{u\sim v} |x_u| \leq \deg(v)|x_v| \leq \Delta|x_v|.$$

This shows $|\lambda_1(A)| \leq \Delta$, proving the result. □

Examining this proof, we see that we hardly used any of the properties of A in our argument. In particular, word for word the same argument gives the following.

Lemma 24.3. *Let G be a graph and M a symmetric matrix such that $M_{i,j} = \pm 1$ if $ij \in E(G)$ and $M_{i,j} = 0$ otherwise. Then*

$$\lambda_1(M) \leq \Delta,$$

where $\lambda_1(M)$ is the largest eigenvalue of M and Δ is the maximum degree of G.

Next we have the Cauchy interlacing theorem, which allows us to bound the eigenvalues of sumbatrices of B in terms of the eigenvalues of B.

Theorem 24.4 (Cauchy interlacing theorem)**.** *Let B be a real symmetric $n \times n$ matrix and C an $m \times m$ principal sumbmatrix of B with $m \leq n$. If B has eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and C has eigenvalues $\mu_1 \geq \cdots \geq \mu_m$, then for all i*

$$\lambda_i \geq \mu_i \geq \lambda_{i+n-m}.$$

Remark 24.5. Personally, I always forget the exact statement of Cauchy's interlacing theorem, so here's a bit of "mnemonic" to help remember it. By the Raleigh quotient we always have

$$\mu_1 = \max_{x\neq 0} \frac{x^*Cx}{x^*x} \leq \max_{x\neq 0} \frac{x^*Bx}{x^*x} = \lambda_1,$$

so that gives the first inequality. We also have $\mu_1 \geq \lambda_1$ if $m = n$, and each time we decreases m this bound has to get weaker, so we end up getting $\mu_1 \geq \lambda_{1+n-m}$ in general. We then get the rest of the inequalities by translating $\lambda_1 \geq \mu_1 \geq \lambda_{1+n-m}$ by $i - 1$.

With these standard spectral graph theory lemmas established, we can now state the main theorem of this subsection.

Theorem 24.6 (Huang [90])**.** *Let Q_n be the hypercube graph on 2^n vertices. If $V \subseteq V(Q_n)$ is a subset of size $2^{n-1}+1$, then the induced subgraph $Q_n[V]$ has maximum degree at least $\sqrt{n}$.*

This result is sharp in several ways. First, it is easy to find subsets of size 2^{n-1} such that $Q_n[V]$ is the empty graph, so in order to get any non-trivial lower bound on the maximum degree one needs V to have size at least $2^{n-1}+1$. Second, Chung et. al. [37] proved that there exist choices of V such that $Q_n[V]$ has maximum degree $\lceil\sqrt{n}\rceil$, so this bound is essentially best possible.

It was shown by Gotsman and Linial [82] that proving a result of this form is equivalent to showing that two notions of "sensitivity" for Boolean functions are equivalent, which led to a great deal of interest in resolving it. Nevertheless, it remained unanswered for 30 years until Huang came up with the following remarkable proof.

The key idea is to define the $2^n \times 2^n$ matrix B_n recursively by

$$B_0 = [0], \qquad B_n = \begin{bmatrix} B_{n-1} & I \\ I & -B_{n-1} \end{bmatrix},$$

where here I denotes the identity matrix of dimension 2^{n-1}. Observe that if the negative sign in the definition of B_n wasn't there, then this would just define the adjacency matrix of Q_n. Thus this is a sort of "twisted adjacency matrix" which has -1's in some of the positions where there are usually 1's. This choice of signings turns out to spread out the spectrum of B_n in a nice way.

Lemma 24.7. *The spectrum $\sigma(B_n)$ consists of $\pm\sqrt{n}$ each occurring with multiplicity 2^{n-1}.*

Proof. It is straightforward to prove by induction that $B_n^2 = nI$, which implies that every eigenvalue λ of B_n satisfies $\lambda^2 = n$. Thus $\sigma(B_n)$ consists of $\pm\sqrt{n}$, and each must appear with equal multiplicity because $\mathrm{Tr}(B_n) = 0$. □

Shockingly, we have everything we need for our proof.

Proof of Theorem 24.6. Let $B = B_n$ be as described above. Let $V \subseteq V(Q_n)$ be any subset of size $2^{n-1} + 1$ and let C be the submatrix of B indexed by the rows and columns corresponding to B. Let $G = Q_n[V]$. Observe that C satisfies the conditions for M of Lemma 24.3 since B is a (symmetrically) signed version of the adjacency matrix. By Lemma 24.3, the Cauchy interlacing theorem, and the previous lemma, we conclude that

$$\Delta(G) \geq \lambda_1(C) \geq \lambda_{2^{n-1}}(B) = \sqrt{n},$$

proving the result. □

24.2 Hoffman's Bound and Erdős-Ko-Rado

One of the most important results in spectral graph theory is Hoffman's bound, which gives an upper bound on the independence number of a regular graph in terms of its eigenvalues.

Lemma 24.8 (Hoffman)**.** *Let G be a non-empty n vertex d-regular graph and A its adjacency matrix. Then*

$$\frac{\alpha(G)}{n} \leq \frac{-\lambda_{\min}}{d - \lambda_{\min}},$$

where $\lambda_{\min}$ is the smallest eigenvalue of A.

Proof. Let I be an independent set of size $\alpha = \alpha(G)$, and let x be the characteristic vector with $x_i = 1$ if $i \in I$ and $x_i = 0$ otherwise. Observe that because I is an independent set, we have

$$x^T A x = 0.$$

Let $v_1, \ldots, v_n$ be an orthonormal eigenbasis for A with eigenvalues $\lambda_1, \ldots, \lambda_n$. Since G is regular, the all 1's vector $\mathbf{1}$ is an eigenvector with eigenvalue equal to d, so we can assume $v_1 = \mathbf{1}/\sqrt{n}$ and $\lambda_1 = d$. Writing $x = \sum c_i v_i$ for some real numbers c_i, we see that

$$\alpha = x^T x = \sum c_i^2,$$

and

$$\alpha/\sqrt{n} = \langle x, v_1 \rangle = c_1.$$

Putting all of this together, we find

$$\begin{aligned} 0 = x^T A x = x^T \sum c_i \lambda_i v_i = \sum c_i^2 \lambda_i = (\alpha^2/n)d + \sum_{i \neq 1} c_i^2 \lambda_i \\ \geq (\alpha^2/n)d + \sum_{i\neq 1} c_i^2 \lambda_{\min} = (\alpha^2/n)d + (\alpha - \alpha^2/n)\lambda_{\min}. \end{aligned}$$

Dividing both sides by α and rearranging gives

$$\alpha(\lambda_{\min} - d)/n \geq \lambda_{\min}.$$

Dividing both sides by $\lambda_{\min} - d$ (which is negative because $\lambda_{\min} < 0$ since G is non-empty) gives the result. □

This result gives yet another proof of Erdős-Ko-Rado.

Sketch of Proof of Erdős-Ko-Rado. Let G be the graph with vertex set $\binom{[n]}{r}$ where two sets are adjacent if and only if they are disjoint. Observe that independent sets of G are exactly intersecting families, so Erdős-Ko-Rado is equivalent to saying that

$$\alpha(G) = \binom{n-1}{r-1}$$

provided $n \geq 2r$. In this case, one can verify that G has eigenvalues $(-1)^j \binom{n-r-j}{r-j}$ for $0 \leq j \leq r$ (each appearing with multiplicity $\binom{n}{j} - \binom{n}{j-1}$), and hence

$$\lambda_{\min} = -\binom{n-r-1}{r-1} = -\frac{r}{n-r}\binom{n-r}{r}.$$

Since G has $\binom{n}{r} = \frac{n}{r}\binom{n-1}{r-1}$ vertices and is regular with degree $\binom{n-r}{r}$, plugging things into Lemma 24.8 gives $\alpha(G) \leq \binom{n-1}{r-1}$ as desired. □

Perhaps the biggest constraint in Lemma 24.8 is the need for G to be regular, which we used in this proof to guarantee that the all 1's vector is an eigenvector, allowing us to extract $\alpha = \langle x, \mathbf{1} \rangle$. We can get around this issue if we change how we measure independent sets. To this end, given a graph G together with a vector w indexed by $V(G)$ and a set of vertices I, define $|I|_w = \sum_{i \in I} w_i^2$, and define $\alpha_w(G) = \max_I |I|_w$ where I ranges over all independent sets of G.

With this we can generalize Lemma 24.8 by both allowing G not to be regular, and by allowing A to only be a "pseudo-adjacency matrix." It will also be important for our main application to allow G to have loops, where we emphasize that no independent set of G can contain a vertex which has a loop.

Lemma 24.9. *Let G be a graph with loops and M a matrix such that $M_{i,j} = 0$ whenever $i \not\sim j$ and such that M has a basis of eigenvectors. If $\lambda_{\min}$ is the smallest eigenvalue of M, and if w is a unit eigenvector of M with eigenvalue $\lambda > \lambda_{\min}$, then*

$$\alpha_w(G) \le \frac{-\lambda_{\min}}{\lambda_{\min} - \lambda}.$$

The proof of Lemma 24.9 is nearly identical to that of Lemma 24.8, and we leave the details as an exercise to the reader.

In order to use Lemma 24.9, we need a problem where we care about independent sets which are weighted in non-standard ways. Motivated by our application of Lemma 24.8, we will do this for a weighted version of the Erdős-Ko-Rado problem.

Specifically, given an n-vertex set system $\mathcal{F}$, we define its p-biased measure by

$$\mu_p(\mathcal{F}) = \sum_{F \in \mathcal{F}} (1-p)^{n-|F|} p^{|F|}.$$

In this setting, our main question is: what is the largest size (in the p-biased sense) of an intersecting set system? Here we emphasize that we consider the non-uniform families, as otherwise this is just equivalent to Erdős-Ko-Rado. The answer to this question is more or less what one would expect (after observing that $\mu_p(\mathcal{F}) = p$ when $\mathcal{F}$ consists of all sets containing a given element).

Theorem 24.10. *If $\mathcal{F}$ is an n-vertex intersecting family and $p \in [0, 1/2]$, then $\mu_p(\mathcal{F}) \le p$.*

One can show that this result is false for $p > 1/2$. It also turns out that for $p < 1/2$ the unique construction with $\mu_p(\mathcal{F}) = p$ is the star, but we'll refrain from going into this.

One can also get uniqueness but I don't know how much extra work this is. Also note that this result is false for larger p

Just like with the original EKR theorem, there exist many proofs of Theorem 24.10 (in fact, Friedgut has a talk going through 5 and a half proofs of this theorem **REF**). The spectral proof we present here is due to Friedgut [69], and this proof ends up having the advantage that a closer analysis of the proof gives stability results.

Mimicking our proof of EKR using Lemma 24.8 described above, we define the graph G_n on $2^{[n]}$ where two sets are adjacent if and only if they are disjoint. Note that G has a loop at the vertex $\emptyset$. Again, independent sets of G_n are intersecting families, and proving Theorem 24.10 is equivalent to showing $\alpha_w(G_n) \le p$ where w is the vector defined by $w_S = (1-p)^{n-|F|} p^{|F|}$ (which is a unit vector).

To complete the proof using Lemma 24.9, it remains to cook up a pseudo-adjacency matrix M_n which in particular has w as an eigenvector. To develop some intuition, we focus on the case $n = 1$. Here G_1 is just an edge where one vertex has a loop, so the M_1 we are looking must have the form $M_1 = \begin{bmatrix} * & * \\ * & 0 \end{bmatrix}$. To limit our search space a little, we will try and find such a matrix which is symmetric (though this isn't required to use Lemma 24.9), at which point we can scale M appropriately so that it has the form

$$M_1 = \begin{bmatrix} c & 1 \\ 1 & 0 \end{bmatrix},$$

for some value c to be determined. Again, we must have that $w = (\sqrt{1-p}, \sqrt{p})$ is an eigenvector for this matrix, which means there exists some λ with $\lambda w_1 = cw_1 + w_2$ and with $\lambda w_2 = w_1$. Plugging things in, we see that $\lambda = \sqrt{(1-p)/p}$, and hence that

$$M_1 = \begin{bmatrix} \sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}} & 1 \\ 1 & 0 \end{bmatrix}$$

We also observe that the other eigenvector of this matrix is $w' = (\sqrt{p}, -\sqrt{1-p})$, which has eigenvalue $\lambda' = -\sqrt{p/(1-p)}$. Plugging this into Lemma 24.9, we find

$$\alpha_p(G) \leq \frac{\sqrt{p/(1-p)}}{\sqrt{(1-p)/p} + \sqrt{p/(1-p)}} = \frac{p}{1-p+p} = p,$$

proving the result when $n = 1$.

It remains to generalize M_1 to work for larger n. While it isn't hard to do this directly, we'll do this in one go with the following linear algebra fact about tensor products.

Lemma 24.11. *Given real symmetric matrices B_1, B_2 with rows and columns indexed by V_1, V_2 respectively, define the tensor product matrix $B_1 \otimes B_2$ by having its rows and columns indexed by $V_1 \times V_2$ with $(B_1 \otimes B_2)_{(i_1,i_2),(j_1,j_2)} = (B_1)_{i_1,j_1}(B_2)_{i_2,j_2}$.*

If $u_1, \ldots, u_{|V_1|}$ and $v_1, \ldots, v_{|V_2|}$ are orthonormal sets of eigenvectors for M_1, M_2 with u_i having eigenvalue μ_i and v_i having eigenvalue λ_i, then the vectors $u_i \otimes v_j$ defined by $(u_i \otimes v_j)_{(k_1,k_2)} = (u_i)_{k_1}(v_j)_{k_2}$ are an orthonormal set of eigenvectors for $M_1 \otimes M_2$ with eigenvalue $\mu_i \lambda_j$.

This is pretty easy to prove once you unwind the definitions. With this in mind, we inductively define $M_n = M_{n-1} \otimes M_1$ (i.e. M_n is the nth tensor power of M_1). Equivalently, this is the matrix defined by $(M_n)_{S,T} = 0$ whenever $S \cap T = \emptyset$ and which has $(M_n)_{S,T} = (\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}})^{n-|S \cup T|}$ otherwise. The point is that by the lemma above, the vector w with $w_S = (1-p)^{(n-|S|)/2} p^{|S|/2}$ is an eigenvector corresponding to the eigenvalue $((1-p)/p)^{n/2}$. Further, if $p \leq 1/2$, then we have

$$\lambda_{\min} = \min_i (-1)^i (p/(1-p))^{i/2} ((1-p)/p)^{(n-i)/2} = -\sqrt{p/(1-p)}) \cdot ((1-p)/p)^{(n-1)/2}.$$

Plugging this into Lemma 24.9 gives

$$\alpha_w(G_n) \leq \frac{\sqrt{p/(1-p)}) \cdot ((1-p)/p)^{(n-1)/2}}{((1-p)/p)^{n/2} + \sqrt{p/(1-p)}) \cdot ((1-p)/p)^{(n-1)/2}} = p,$$

proving Theorem 24.10.

Comment on uniqueness and stability; maybe go through this for $t = 1$.

With some work, one can extend the approach of this subsection to work for t-intersecting families. Unfortunately the most naive approach of solving the problem for $n = 1$ and then tensorizing doesn't work (since the $n = 1$ case is too small to capture the situation for $t > 1$). The way Friedgut gets around this in [69] is by doing linear algebra over the ring $\mathbb{R}[X]/(X^t = 0)$ instead of just $\mathbb{R}$. In this setting, he roughly considers matrices where the S, T entry is equal to $c_{S,T}X^{|S\cap T|}$ for some real number $c_{S,T}$, which makes it so that these entries are 0 whenever $|S \cap T| \geq t$. A lot of the steps become more intricate in this setup, but eventually everything does end up going through.

25 The Polynomial Method I: Slice Rank

Roughly speaking, the *polynomial method* is the method of using polynomials to solve problems, though there is no real agreed upon definition of what does and does not constitute an instance of this method.

One might argue that the modular intersection section used this method, in that we obtained our results by cooking up a system of (low degree) polynomials which "encoded" our problem (i.e. by choosing them so that the roots of the polynomials corresponded to intersections being of a given size). Over the next two sections, we will look at two other systematic ways one can take (low degree) polynomials which "encode" your problem and translate them into bounds for an extremal problem: slice rank and combinatorial Nullstellensatz. We note that these two methods are independent of each other and can be read in any order.

25.1 Slice Rank

In the previous section we considered proofs using linear algebra. Here we go a step further and use *multilinear* algebra.

To this end, given a set S and an integer k, we let S^k denote the set of all k-tuples of elements of S. Given a field $\mathbb{F}$, we will say that any function of the form $f : S^k \to \mathbb{F}$ is a *k-tensor*. Note that when $k = 2$, we can express f as a matrix whose rows and columns are indexed by S, and as such we can think of k-tensors as "higher order" matrices. Our main goal here is to use "ranks" of tensors to obtain bounds for combinatorial problems, analogous to what we did in our second proof of Oddtown.

There are various non-equivalent ways one can generalize the notion of rank from matrices to tensors. One way is through *slice rank*. For this, we say that a k-tensor f is a *slice* if

$$f(x_1, \ldots, x_k) = g(x_i)h(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k),$$

where i is some integer, g is a 1-tensor, and h is a $(k-1)$-tensor. Note that when $k = 2$, slices are exactly rank 1 matrices, i.e. the outer product of two vectors g and h. We define the slice rank $sr(f)$ of a k-tensor f to be the smallest integer r such that f can be written as the sum of r slices. Again note that in the case $k = 2$ this exactly corresponds to the usual notion of rank.

Although not every property of matrix rank carries over to the setting of slice rank, one important property that does is the fact that the slice rank of a "diagonal tensor" equals the number of non-zero entries it has. To this end, we say that a tensor is *proper diagonal* if $f(\vec{x}) \neq 0$ if and only if $x_i = x_j$ for all i, j.

Lemma 25.1. *If $f : S^k \to \mathbb{F}$ is a proper diagonal k-tensor, then $sr(f) = |S|$.*

Proof Sketch. We first show $sr(f) \leq |S|$. For each $a \in S$, let $g_a(x)$ be the 1-tensor with $g_a(a) = 1$ and $g_a(b) = 0$ otherwise. We observe

$$f(\vec{x}) = \sum_{a \in S} g_a(x_1) f(a, x_2, \ldots, x_k),$$

proving that f has slice rank at most $|S|$ as desired.

The proof of the lower bound $sr(f) \geq |S|$ is a somewhat fiddly induction argument that takes about two pages to do; we refer the reader to a nice writeup by Martinez [112, Lemma 2.2.9] for the full details. $\square$

Our general approach using Lemma 25.1 will be as follows:

(1) Start with some set S which has some desired properties.

(2) Using the properties of S, we construct a proper diagonal k-tensor $f : S^k \to \mathbb{F}$ which is "simple" (e.g. a polynomial of low degree).

(3) Using the fact that f is "simple", we find some explicit way to write f as the sum of r slices.

(4) In total, steps (2) and (3) imply

$$|S| = sr(f) \leq r,$$

giving us an effective bound on sets S with our desired properties.

We note that this sort of approach was first used by Croot, Lev, and Pach [44]. This method was then adapted by Ellenberg and Gijswitj [54] before being systemetized by Tao [140].

We will use the general approach outlined above to give very short proofs of two problems which were previously thought to be incredibly difficult. Throughout this section, we suggest before the reader goes through each proof that they first sit down and try to construct an f as in the framework outlined above in order to solve the problem.

25.2 The Capset Problem

Given an abelian group G, we say that a triple of elements $(x, y, z) \in G^3$ forms a 3-term arithmetic progression (or 3AP for short) if $y - x = z - y$, and we say this is a non-trivial 3AP if we do not have $x = y = z$. A fundamental question in additive combinatorics is to determine how large a subset $S \subseteq G$ can be if it contains no non-trivial 3AP. For example, Roth's theorem famously says that when $G = \mathbb{Z}$ such a set must have density 0.

Another natural case to consider is $\mathbb{F}_p^n$ for prime p. The case $p = 2$ is trivial (if $S \subseteq \mathbb{F}_2^n$ contains two distinct elements x, y, then (x, y, x) forms a non-trivial 3AP). As such, the first interesting case is to look at $\mathbb{F}_3^n$, and in this setting sets $S \subseteq \mathbb{F}_3^n$ without non-trivial 3AP's are referred to as cap sets. Using a difficult Fourier analytic argument, Bateman and Katz [17] showed that capsets can have size at most $O(3^n/n^{1+\epsilon})$ and Edel [53] gave the best known lower bound of about 2.2^n. At one point Tao mentioned that this capset problem was perhaps his favorite open problem and that he thought the answer was probably $(3 - o(1))^n$. It was thus a major shock when Ellenberg and Gijswijt [54] gave a remarkably short proof showing an upper bound of $(3 - \epsilon)^n$.

Theorem 25.2 (Ellenberg-Gijswijt [54]). *For all primes p, there exists a real number $\epsilon > 0$ such that any set $S \subseteq \mathbb{F}_p^n$ which contains no non-trivial 3AP has $|S| \leq (p-\epsilon)^n$.*

Again, the reader is encouraged to try and come up with a tensor f which might work to give the result before going through the details of this proof.

Proof. For slight ease of notation we will only deal with the case $p = 3$, the proof for general p being almost identical. As outlined above, our goal will be to use S to construct a proper diagonal tensor f of low degree. Given the inputs of the problem, perhaps the most natural kind of tensor to consider is one of the form $f : S^3 \to \mathbb{F}_3$, where again we want to ensure that if $x, y, z \in S$ then $f(x, y, z) \neq 0$ if and only if $x = y = z$. Critically, because of our hypothesis that S contains no non-trivial 3AP, we see that we want $f(x, y, z) \neq 0$ if and only if x, y, z is a 3AP, and as such our goal is essentially equivalent to constructing an f which is the indicator function for 3AP's!

By definition, (x, y, z) is a 3AP if and only if $y - x = z - y$, and rearranging gives $x - 2y + z = 0$. As such, we're essentially left with the problem of constructing an indicator function for $x - 2y + z \neq 0$, or equivalently that $x_i - 2y_i + z_i \neq 0$ for all i. And this is somewhat easy: simply take

$$f(x, y, z) = \prod_{i=1}^{n}(1 - (x_i - 2y_i + z_i)^2),$$

which one can readily check is an indicator function for 3AP's in S, and hence is a proper diagonal tensor.

It remains to estimate the slice rank of f. That is, we want to show that f can be written as the sum of a small number of functions of the form e.g. $g(x)h(y, z)$. To this end, we make the observation that by the definition of f given above, each of its monomials has degree at most $2n$, and as such each of its monomials has one of its x-degree, y-degree, or z-degree is at most $2n/3$. The idea now is to group each of these monomials with the same low degree variable and then bound the slice rank of each of these groups by using the low degree variable.

To be somewhat more precise, given a vector $\alpha \in \{0, 1, 2\}^n$, define $x^\alpha = \prod x_i^{\alpha_i}$, and similarly define y^β and z^γ. Let $|\alpha| := \sum \alpha_i$, and let $\mathcal{M}$ denote the set of all triples (α, β, γ) with $\alpha, \beta, \gamma \in \{0, 1, 2\}^n$ and $|\alpha| + |\beta| + |\gamma| \leq 2n$. Observe that each monomial of f is of the form $x^\alpha y^\beta z^\gamma$ for some $(\alpha, \beta, \gamma) \in \mathcal{M}$.

Let $\mathcal{M}_x \subseteq \mathcal{M}$ be the set of triples with $|\alpha| \leq 2n/3$, let $\mathcal{M}_y \subseteq \mathcal{M} \setminus \mathcal{M}_x$ be those with $|\beta| \leq 2n/3$, and let $\mathcal{M}_z \subseteq \mathcal{M} \setminus (\mathcal{M}_x \cup \mathcal{M}_y)$ be those with $|\gamma| \leq 2n/3$. As noted above, $\mathcal{M}_x \cup \mathcal{M}_y \cup \mathcal{M}_z$ partition $\mathcal{M}$. Given α with $|\alpha| \leq 2n/3$, define the polynomial

$$h_\alpha(y, z) = \sum_{\beta,\gamma:(\alpha,\beta,\gamma)\in\mathcal{M}_x} c_{\alpha,\beta,\gamma} y^\beta z^\gamma,$$

where $c_{\alpha,\beta,\gamma}$ is the (possibly 0) coefficient of $x^\alpha y^\beta z^\gamma$ in f. Similarly define $h_\beta(x, z)$ and $h_\gamma(x, y)$. By definition and the fact that $\mathcal{M}_x \cup \mathcal{M}_y \cup \mathcal{M}_z$ partition $\mathcal{M}$, we have

$$f(x, y, z) = \sum_{\alpha:|\alpha|\leq 2n/3} x^\alpha h_\alpha(y, z) + \sum_{\beta:|\beta|\leq 2n/3} y^\beta h_\beta(x, z) + \sum_{\gamma:|\gamma|\leq 2n/3} z^\gamma h_\gamma(x, y).$$

With this, we conclude that the slice rank of f is at most 3 times the number of terms in eahc sum, which is equal to the number of vectors $\alpha \in \{0,1,2\}^n$ with $|\alpha| \leq 2n/3$. It is not convince oneself that the number of such vectors is at most $(3-\epsilon)^n$ for some ϵ (essentially because a random $\alpha \in \{0,1,2\}^n$ behaves like the sum of two bionomial random variables with total expectation n), proving the result. □

We note in particular in the case $p = 3$ one can go through the analysis and get an upper bound of roughly $|S| = O(2.756^n)$ for sets $S \subseteq \mathbb{F}_3^n$ not containing a capset. In fact, it turns out that this upper bound continues to hold for a slightly more general problem: we say that three sets $X, Y, Z \subseteq \mathbb{F}_3^n$ contain no rainbow 3AP if the only 3AP's $(x, y, z) \in X \times Y \times Z$ are trivial. With a bit more work one can show that the above proof yields that if X, Y, Z contains no rainbow 3AP then $|X|+|Y|+|Z| = O(2.756^n)$, and moreover there exist constructions that essentially match this upper bound. Thus any improvement to the bound for capsets has to somehow utilize that we only allow triples from S^3 and not from three distinct sets. We refer to the interested reader to the nice survey article by Grochow [84] for more on this topic.

25.3 Nonuniform Sunflowers

Recall that a k-sunflower is a collection of sets $h_1, \ldots, h_k$ such that $h_i \cap h_j$ is equal to the same set for all $i \neq j$. In the setting of uniform hypergraphs, the famous Erdős-Rado sunflower conjecture says that there exists a constant $C = C(k)$ such that every r-uniform hypergraph with at least C^r hyperedges contains a k-sunflower.

Here we consider an analogous conjecture in the non-uniform setting due to Erdős and Szemerédi [61]: there exists a constant $\epsilon = \epsilon(k) > 0$ such that for all $n \geq \epsilon^{-1}$, every $\mathcal{H} \subseteq 2^{[n]}$ with at least $(2-\epsilon)^n$ hyperedges contains a k-sunflower. This conjecture turns out to be weaker than the Erdős-Rado sunflower conjecture [61].

The slice rank method allows us to give an easy proof of this non-uniform sunflower conjecture when $k = 3$.

Theorem 25.3 (Naslund-Sawin [122])**.** *If $\mathcal{H}$ is an n-vertex hypergraph without any 3-sunflower, then*

$$|\mathcal{H}| \leq (n+1) \sum_{t \leq n/3} \binom{n}{t} \leq 1.89^n.$$

Proof. Let S be the set of characteristic vectors corresponding to the hyperedges of $\mathcal{H}$. As a first attempt for this problem, one would probably try to work with the set S directly to construct a tensor f, but this causes some technical issues pop up.

To get around these issues, we let $S_r \subseteq S$ be those characteristic vectors with r 1's. The insight here is that $\max_r |S_r| \leq |S| \leq (n+1) \cdot \max_r |S_r|$, so up to some negligible factor it suffices to bound the size of each S_r set. This will allow us to get around the previously mentioned issues.

With this, our goal is to construct a 3-tensor f, now with domain S_r^3, such that $f(x, y, z) \neq 0$ if and only if $x = y = z$. The crucial insight here is that if we do not have $x = y = z$, then there exists some i such that $x_i + y_i + z_i = 2$. Indeed, if this were not the case and x, y, z were all pairwise distinct from each other, then this would imply that every i is either contained in 0, 1, or 3 of the sets x, y, z, which would imply these distinct hyperedges form a 3-sunflower. If, say $x = y \neq z$, then the non-existence of such an i would imply $x \subseteq z$, a contradiction to $x \neq z$ both being in S_r, i.e. both having the same number of elements.

With this observation in mind, we see that our desired f is essentially the indicator function for [no coordinate i having $x_i + y_i + z_i = 2$]. Thus we can take $f : S_r^3 \to \mathbb{F}_3$ by defining

$$f(x, y, z) = \prod_{i=1}^{n} (x_i + y_i + z_i - 2),$$

and one can easily check that f is a proper diagonal tensor given that $S_r \subseteq S$ contains no 3-sunflowers, so it remains to bound the slice rank of f. Similar to before, we observe that each monomial of f has either x, y or z degree at most $n/3$, so by a similar trick as before we conclude

$$\max_r |S_r| \leq sr(f) \leq \sum_{t \leq n/3} \binom{n}{t} \leq 2^{nH(1/3)},$$

where $H(p) = -p \log_2(p) - (1-p)\log_2(1-p)$ is the binary entropy function. Evaluating this (and multiplying by $n+1$) gives the desired upper bound. $\square$

26 The Polynomial Method II: Combinatorial Nullstellensatz

Roughly speaking, the main idea of the previous section was to construct low degree polynomials f that were capable of encoding any set S with a given property; and from there we used slice rank to show that any such set S must be small in terms of $\deg(f)$.

In this subsection we take a subtlely different approach: we construct polynomials f in terms of some given set S (generally with degree $|S|$), and then use the properties of S to argue that f must have relatively small degrees.

To emphasize, in the previous subsection we constructed f which was independent of the specific S we were given with some property, but here we construct f which will defined based off of whatever S is given to us. This is somewhat more reminiscent of the approach we took with modular intersections.

In any case, the main tool we will use for this new approach is the following result of Alon's, which is inspired by Hillbert's Nullstellensatz theorem from algebraic geometry. Roughly speaking, it says that if f has small degree, then f can not vanish on a large cartesian product.

Theorem 26.1 (Alon's Combinatorial Nullstellensatz [1])**.** *Let $\mathbb{F}$ be a field and $f \in \mathbb{F}[x_1, \ldots, x_n]$ a polynomial. If $t_1, \ldots, t_n$ are non-negative integers such that $\deg(f) = \sum t_i > 0$ and such that the monomial $\prod x_i^{t_i}$ has a non-zero coefficient in f, then for any sets $S_i \subseteq \mathbb{F}$ satisfying $|S_i| > t_i$ for all i, there exist elements $s_i \in S_i$ such that $f(s_1, \ldots, s_n) \neq 0$.*

We will somewhat informally write the conclusion of this statement as $f(S_1, \ldots, S_n) \neq 0$ (and we will also write $f(S_1, \ldots, S_n) = 0$ whenever this conclusion fails to hold). We note that this result is best possible, as can be seen by taking the function $\prod_{s \in S_1}(x_1 - s)g(x_2, \ldots, x_n)$ for any set $S_1 \subseteq \mathbb{F}$ and function g.

Proof. We give a short proof due to Michałek [117] by induction on $\deg(f)$, the case $\deg(f) = 1$ being straightforward. Let f, $t_1, \ldots, t_n$, and $S_1, \ldots, S_n$ be as in the hypothesis, and assume we have proven the result for all polynomials of degree smaller than $\deg(f) > 1$. Without loss of generality, we can assume $t_1 > 1$, and we let $s \in S_1$ be an arbitrary element. Using polynomial division, we can write

$$f(x_1, \ldots, x_n) = (x_1 - s)g(x_1, \ldots, x_n) + h(x_1, \ldots, x_n)$$

such that $\deg(g) = \deg(f) - 1$ and such that h does not depend on x_1. If we assume for contradiction that $f(S_1 \times \cdots \times S_n) = 0$, then in particular we have

$$0 = f(\{s\} \times S_2 \times \cdots \times S_n) = 0 + h(\{s\} \times S_2 \times \cdots \times S_n).$$

Because h does not depend on x_1, this implies $h(S_1 \times \cdots \times S_n) = 0$. Again using that h does not depend on x_1 and the hypothesis of the theorem, g must have a monomial $x_1^{t_1 - 1} \prod_{i \neq 1} x_i^{t_i}$ with a non-zero coefficient, so by induction we can find some $(s_1, \ldots, s_n) \in (S_1 \setminus \{s\}) \times S_2 \times \cdots \times S_n$ which g does not vanish on. Note that $h(s_1, \ldots, s_n) = 0$ by our previous observation, so in total we conclude that $f(s_1, \ldots, s_n) = (s_1 - s)g(s_1, \ldots, s_n) \neq 0$, a contradiction. $\square$

We will now showcase a slew of examples which use the combinatorial Nullstellensatz, with many more examples being found in Alon's original paper [1]. We note that in most cases, proofs of these results were known well before the introduction of the combinatorial Nullstellensatz, but this tool allows for elegant and unified solutions to all of them. Again, the reader is encouraged to try and figure out which polynomial to use before reading through the proof.

26.1 Cauchy-Davenport

We begin with a basic result from additive combinatorics. Given two subsets A, B of an abelian group G, we define $A + B = \{a + b : a \in A,\ b \in B\}$. If $G = \mathbb{Z}$, then it is not difficult to show that $|A+B| \geq |A| + |B| - 1$, which is best possible. However, if G is a finite group, then this bound no longer holds for the simple reason that one could have $|A| + |B| - 1 > |G|$. The Cauchy-Davenport theorem says that this is the only obstruction for $\mathbb{F}_p$.

Proposition 26.2 (Cauchy-Davenport)**.** *If $A, B \subseteq \mathbb{F}_p$ for some prime number p, then $|A+B| \geq \min\{|A| + |B| - 1, p\}$.*

Proof. Let's think for a moment about what parameters we might choose when utilizing Theorem 26.1. Seemingly we should take $\mathbb{F} = \mathbb{F}_p$. Since we ultimately want to conclude something about the size of $A + B$, it seems reasonable that we want to construct our polynomial f to depend on $A + B$, and in particular to have degree equal to $|A + B|$. The only relevant sets we have that could play the role of the S_i are $A, B, A + B$ and $\mathbb{F}_p$. Since we're already planning to use $A + B$ to construct our polynomial, it perhaps makes the most sense to try and use $S_1 = A$ and $S_2 = B$.

Roughly then, we want to construct a polynomial $f(x, y)$ of degree equal to $|A + B|$, which means that if $|A+B| = \deg(f) \leq (|A|-1)+(|B|-1)$, then by the combinatorial Nullstellensatz we will have $f(A, B) \neq 0$. We want to then derive a contradiction from this and from this; that is, we want to define f in such a way that we obviously have $f(A, B) = 0$.

With the discussion above in mind, we want to define an f such that $f(A, B) = 0$ and such that f has degree $|A + B|$. An obvious candidate for this is to take

$$f(x, y) = \prod_{c \in A+B} (x + y - c),$$

which has all of the properties described above.

To complete the proof, we assume for contradiction that $|A+B| \leq \min\{|A|+|B|-2, p-1\}$. To apply the Nullstellensatz with $S_1 = A,\ S_2 = B$, we need to show that there exists a monomial $x^{t_1} y^{t_2}$ with $t_1 + t_2 = |A + B|$, $t_1 < |A|$, and $t_2 < |B|$. To this end, consider $t_1 = |A| - 1$ and $t_2 = |A + B| - |A| + 1 \leq |B| - 1$, with this inequality coming from our assumption on $A + B$. It is not difficult to see that the coefficient of $x^{t_1} y^{t_2}$ in f is equal to $\binom{|A+B|}{|A|-1}$. Crucially, because $|A + B| < p$, this binomial coefficient is not a multiple of p and hence is non-zero. We can thus apply the Nullstellensatz with $S_1 = A,\ S_2 = B$ to conclude that $f(A, B) \neq 0$, a contradiction to how we defined f. $\square$

26.2 Covering the Hypercube

How many hyperplanes in $\mathbb{R}^n$ does it take to cover all of the points of the hypercube $\{0,1\}^n$? It's not difficult to see that the minimum you need is 2. To make things more interesting, we ask the following variant for affine hyperplanes: how many hyperplanes do you need to cover all of $\{0,1\}^n \setminus (0,\ldots,0)$ such that none of the hyperplanes contain $(0,\ldots,0)$? It isn't too difficult to work out that n hyperplanes suffice, but proving this is best possible isn't as easy. However, this becomes doable with the Nullstellensatz.

Theorem 26.3 (Alon-Füredi [5]). *If $h_1,\ldots,h_m$ are a sequence of hyperplanes in $\mathbb{R}^n$ which do not contain the origin and which contain every other point of $\{0,1\}^n$, then $m \geq n$.*

Proof. Again let us discuss what sort of parameters we might choose in applying the Nullstellensatz. The most natural choice for field is probably $\mathbb{R}$. Again our polynomial f should probably have degree m, and a natural way to do this is to take f to be a product over terms indexed by the h_j. Motivated by this, for each j let $a_j \in \mathbb{R}^n$ and $b_j \in \mathbb{R}$ be such that h_j contains all of the points $x \in \mathbb{R}^n$ satisfying $\langle x, a_j\rangle = b_j$, and we note for later that $b_j \neq 0$ for all j since the origin is not contained in any h_j. Then a natural choice of $f : \mathbb{R}^n \to \mathbb{R}$ to consider is

$$f(x) = \prod_{i=1}^{m} (b_j - \langle x, a_j\rangle),$$

since by hypothesis we know that our choice of f vanishes on all of $\{0,1\}^n$ except the origin.

At this point, if we assume for contradiction that $m \leq n-1$, then we know f has degree at most $n-1$, and hence there's a chance we can find a monomial of the form $\prod_{i\neq k} x_i$ for some k with a non-zero coefficient. In this case we could apply the Nullstellensatz with $S_i = \{0,1\}$ for $i \neq k$ and $S_k = \{1\}$ to find a point which isn't the origin that f fails to vanish on, a contradiction to how we defined f.

While the above approach is theoretically possible, it's not at all clear how you could correctly choose the value of k given the very limited information of the problem setup. Thus instead of doing the above, we'll modify our polynomial f so that it vanishes at every point of $\{0,1\}^n$ (including the origin). Since f only fails to vanish at the origin, this can easily be done by taking

$$g(x) = \prod_{j=1}^{m} (b_j - \langle x, a_j\rangle) - \prod_{j=1}^{m} b_j \cdot \prod_{i=1}^{n} (1 - x_i),$$

and it is striaghtforward to check that g now vanishes on all of $\{0,1\}^n$. Moreover, because $f(x)$ had degree $m \leq n-1$ and because $\prod b_j \neq 0$ (as noted at the start of the proof); we have that $g(x)$ has degree equal to n with its only monomial of this degree equal to $\prod x_i$. Thus we can apply the Nullstellensatz with $S_i = \{0,1\}$ for all i to conclude that $g(\{0,1\}^n) \neq 0$, a contradiction to how we constructed g. We conclude that we must have $m \geq n$ as desired. $\square$

We note that this proof features a common trick with the Nullstellensatz: often you will have some "main term" in your polynomial which encodes the bulk of your problem, and from there you add in some additional terms to forbid certain "bad" witnesses $s_1,\ldots,s_n$.

26.3 Regular Subgraphs mod p

Our last example concerns finding subgraphs of graphs which are "regular mod p". The proof will showcases a useful trick one can utilize together with the Nullstellensatz: over the finite field $\mathbb{F}_p$, the function $1 - x^{p-1}$ is the indicator function for x being equal to 0 (since $x^{p-1} = 1$ whenever $x \neq 0$).

Proposition 26.4. *Let p be a prime. If G is an n-vertex graph with more than $(p-1)n$ edges, then G has a non-empty subgraph H where every degree is a multiple of p.*

Proof. Again lets consider the parameters. We should take $\mathbb{F} = \mathbb{F}_p$. One might first consider taking $f : \mathbb{F}_p^n \to \mathbb{F}_p$ with our variables indexed by vertices of G, but in fact the more appropriate domain is $\mathbb{F}_p^{e(G)}$ where each variable x_e corresponds to an edge $e \in G$. This is because we ultimately want to construct a subgraph of G, which is most naturally thought of as a subset of the edges. Since we want our edges to either be in or out of this subgraph, it seems most natural to take $S_e = \{0, 1\}$, where we'll think of $x_e = 1$ meaning $e \in H$. As in the proof of Alon-Füredi, we'll need to ensure $f(0) = 0$ so that H will be non-empty.

Given this setup, we need to cook up an f which encodes the degrees of our vertices being a multiple of p in H. The degree of a vertex v in H is exactly equal to $\sum_{e \ni v} x_e$, and as noted before the proof, we can turn this into an indicator function for being a multiple of p by taking $1 - (\sum_{e \ni v} x_e)^{p-1}$. We again want to make sure we vanish at $x = 0$, so in total the function we come to is

$$f(x) = \prod_v \left(1 - \left(\sum_{e \ni v} x_e\right)^{p-1}\right) - \prod_e (1 - x_e^{p-1}).$$

Note that the degree of the left term is $(p-1)n$ while the degree on the right is $e(G) > (p-1)n$. Thus f is a polynomial of degree $e(G)$ with $\prod x_e^{p-1}$ a monomial achieving this. Thus we can apply the Nullstellensatz with $S_e = \{0, 1\}$ for all e to conclude there's some $x \in \{0, 1\}^n$ on which f does not vanish. By construction we know $x \neq 0$, and hence the graph $H = \{e : x_e \neq 0\}$ is a non-empty subgraph of G, which by construction has every vertex a multiple of p (since if some vertex failed to have this, then we would have $f(x) = 0$). $\square$

Part VIII

Other Methods

27 Trees in Expanders

This section introduces variants of the "extendability method", which are techniques aimed at finding trees in graphs with certain expansion properties. We begin with two relatively simple approaches based on work of Friedman and Pippenger, after which we informally discuss more advanced versions of these ideas before closing with some very brief sketches of how these methods are used in practice.

Throughout this chapter, we let $N_G(X)$ denote the set of vertices in a graph G which have a neighbor in the set $X \subseteq V(G)$. For notational convenience, we will sometimes denote the vertex set of a graph S simply by S rather than $V(S)$ whenever this is clear from context.

27.1 The Friedman-Pippenger Theorem

We say that a graph G is (d, m)-expanding if $|V(G)| > 0$ and if every $X \subseteq V(G)$ with $|X| \leq m$ has $|N_G(X)| \geq d|X|$. We begin by proving the following nice result of Friedman and Pippenger.

Theorem 27.1 ([70]). *If G is $(d+1, 2m)$-expanding, then G contains a copy of every tree T with $m+1$ vertices and maximum degree at most d.*

The idea of the proof is to start with a subgraph $S \subseteq G$ consisting of a single vertex of G, and then to iteratively add leaves to S until we build a copy of T. It is not difficult to convince oneself that one has to be careful with how one adds these leaves to S, as otherwise we might find ourselves with some S which we are not able to extend with a new leaf. As such, we will want to maintain that the S subgraphs are "nice" in some suitable sense so that we never end up stuck.

The formal definition of what it means for S to be "nice" is a little technical, so we will start by trying to develop some intuition for this definition. Eager readers are welcome to skip straight to the formal definition statement at the start of Section 27.1.2.

27.1.1 Intuition

As noted above, we want to prove the Friedman-Pippenger theorem by constructing subgraphs $S \subseteq G$ such that, whenever S is sufficiently small, we can add a new leaf to every vertex $x \in S$. However, this may not always be possible. In particular, to add a leaf to some $x \in S$, we must in particular have $|N_G(x) \setminus S| > 0$, which may not hold if we've chosen S poorly.

With the above in mind, we see that we must build our S with the property that $N_G(x) \setminus S$ is sufficiently large. In particular, since we eventually want to turn S into a copy of T which contains vertices of degree at most d, we should be safe if we impose the condition $|N_G(x) \setminus S| \geq d - \deg_S(x)$ (since in particular, the latter quantity is the maximum number of leaves we'll try adding to x). Somewhat more generally, we will want to ensure that for any $X \subseteq V(S)$, we have that $|N_G(X) \setminus S| \geq \sum_{x \in X} d - \deg_S(x)$, since again we may need to add this many leaves to the set X at some point.

The above only concerns neighborhood conditions for vertices in S, but this is too weak of a condition to impose. Indeed, even if we know some $x \in S$ is such that x has many neighbors in

$G \setminus S$, we may not actually be able to use some of these neighbors $y \in N_G(x) \setminus S$ in any extension of S if adding y would violate the neighborhood condition above. As such, it is essential that we not only guarantee that $|N_G(X) \setminus S|$ is large for any subset $X \subseteq V(S)$, but more generally that this is large for any subset $X \subseteq V(G)$ which could plausibly be used in any extension of S. In particular, we will want to ensure that

$$|N_G(X) \setminus S| \geq d|X| - \sum_{x \in X \cap V(S)} \deg_S(x)$$

for all "small" sets X. The question now is: what do we mean by small here?

Naively, the correct notion of what it means for a set X to be small should be that $|X| \leq m+1$ (since our target subgraph T only has $m+1$ vertices). However, during our proof we will need to have some additional control over the unions $X \cup Y$ where $|X|, |Y| \leq m$. As such, our notion of small sets will encompass all sets of size at most $2m$.

27.1.2 Proof of Theorem 27.1

With the intuition above in mind, we make the following key definition.

Definition 3. Given a graph G, we say that a subgraph $S \subseteq G$ is (d, m)-good if S has at most $m+1$ vertices and maximum degree at most d, and if for all $X \subseteq V(G)$ with $|X| \leq 2m$, we have

$$C(X; S) := |N_G(X) \setminus S| - d|X| + \sum_{x \in X \cap V(S)} \deg_S(x) \geq 0.$$

For example, we claim that if G is $(d+1, 2m)$-expanding for some $d, m \geq 1$, then every vertex $x \in V(G)$ is a (d, m)-good subgraph. Indeed, the size and degree condition for $S = \{x\}$ is automatically satisfied. On the other hand, because G is $(d+1, 2m)$-expanding, we have that $|N_G(X) \setminus S| \geq (d+1)|X| - 1$ for all non-empty X with $|X| \leq 2m$. From this it quickly follows that all such X have $C(X; S) \geq 0$, proving the claim.

The key property we need to show regarding (d, m)-good subgraphs is the following.

Lemma 27.2. *If G is a $(d+1, 2m)$-expanding graph and $S \subseteq G$ is a (d, m)-good subgraph such that $|V(S)| \leq m$ and such that there exists some $v \in S$ with $\deg_S(v) \leq d-1$, then there exists a (d, m)-good subgraph $S \subseteq S' \subseteq G$ obtained by adding a leaf to v.*

Observe that this result, together with induction and the claim above showing that every vertex of G is (d, m)-good, will suffice to prove Theorem 27.1. As such, it remains to prove this lemma. For this, it will be useful to get a better understanding of sets X which have $C(X; S) = 0$ (i.e. which are very close to showing that S is not good). This is accomplished by the following technical result, the proof of which the reader may wish to postpone in order to see how it is used to prove Lemma 27.2.

Proposition 27.3. *Let G, S be as in Lemma 27.2. If X, Y are sets with $C(X; S), C(Y; S) = 0$ and $|X|, |Y| \leq 2m$, then we in fact have $|X|, |Y| \leq m$ and $C(X \cup Y; S) = 0$.*

Proof. We first establish the size condition.

Claim 27.4. *We have $|X|, |Y| \leq m$.*

Proof. Observe that

$$\begin{aligned} 0 = C(X; S) &= |N_G(X) \setminus S| - d|X| + \sum_{x \in X \cap V(S)} \deg_S(x) \\ &\geq |N_G(X)| - |V(S)| - d|X| \\ &\geq (d+1)|X| - m - d|X|, \end{aligned}$$

where the second inequality used that $|V(S)| \leq m$, and that G is $(d+1, 2m)$-expanding and $|X| \leq 2m$. Rearranging gives the desired result for $|X|$, with an identical result giving the bound for $|Y|$. □

We next establish that C behaves nicely with respect to unions (and intersections).

Claim 27.5. *The function $C(\cdot; S)$ is submodular, i.e. for any sets A, B we have $C(A \cap B; S) + C(A \cup B; S) \leq C(A; S) + C(B; S)$.*

Proof. It is straightforward to check that the last two terms $-d|\cdot| + \sum_{x \in \cdot \cap V(S)} \deg_S(x)$ in the definition of $C(\cdot; S)$ form a modular function (i.e. the desired inequality holds with equality). Similarly one can check that $|N_G(\cdot) \setminus S|$ is submodular due to the relations $N_G(X \cup Y) = N_G(X) \cup N_G(Y)$ and $N_G(X \cap Y) \subseteq N_G(X) \cap N_G(Y)$, so combining these two gives a submodular function. □

Because S is (d, m)-good, we have that $C(X \cap Y; S) \geq 0$, and hence the previous claim implies that

$$C(X \cup Y; S) \leq C(X \cap Y; S) + C(X \cup Y; S) \leq C(X; S) + C(Y; S) = 0.$$

On the other hand, Claim 27.4 implies that $|X \cup Y| \leq 2m$, so S being (d, m)-good implies $C(X \cup Y; S) \geq 0$, finishing the proof. □

We now prove our main lemma.

Proof of Lemma 27.2. Let G, S, v be as in the lemma statement. Let $L := N_G(v) \setminus S$, and for each $\ell \in L$, define S_ℓ to be the subgraph of G obtained by adding the edge $v\ell$ to S. The result will follow if S_ℓ is (d, m)-good for some $\ell \in L$, so we may assume for contradiction that this is not the case. As each S_ℓ has at most $m+1$ vertices and maximum degree at most d by hypothesis, these can only fail to be (d, m)-good if for each $\ell \in L$ there exists some set X_ℓ of size at most $2m$ satisfying $C(X_\ell; S_\ell) < 0$.

Because S was (d, m)-good, we have $C(X_\ell; S) \geq 0$ for all $\ell \in L$, and we aim to use this together with $C(X_\ell; S_\ell) < 0$ to derive some structural information about X_ℓ. Specifically, letting $\mathbb{1}[a \in A] = 1$ if $a \in A$ and $\mathbb{1}[a \in A] = 0$ otherwise, it is not difficult to check that

$$C(X_\ell; S_\ell) = C(X_\ell; S) - \mathbb{1}[\ell \in N_G(X_\ell)] + \mathbb{1}[v \in X_\ell] + \mathbb{1}[\ell \in X_\ell],$$

since having $\ell \in N_G(X_\ell)$ would cause the $|N_G(X_\ell) \setminus S|$ term to decrease by 1 as we add ℓ to S, and similarly either $v \in X_\ell$ or $\ell \in X_\ell$ would cause the $\sum \deg_S(x)$ term to increase by 1 as we go to S_ℓ. The only way then that we can have $C(X_\ell; S_\ell) < 0$ and $C(X_\ell; S) \geq 0$ is is if (1) $\ell \in N_G(X_\ell)$, (2) $v \notin X_\ell$, (3) $\ell \notin X_\ell$, and (4) $C(X_\ell; S) = 0$.

Let X' denote the union of all of the X_ℓ sets. By (4) and Proposition 27.3, we see that $C(X'; S_\ell) = 0$ and hence $|X'| \leq s - 1$. Let $X'' = X' \cup \{v\}$. Using $X' \subseteq X''$, the definition of L, and (1), we find that

$$(N_G(X') \setminus S) \subseteq (N_G(X'') \setminus S) = (N_G(X') \setminus S) \cup (N_G(v) \setminus S) = (N_G(X') \setminus S) \cup L = (N_G(X') \setminus S),$$

where we emphasize this last step used that (1) implies $L \subseteq N_G(X') \setminus S$. We thus conclude that $N_G(X'') \setminus S = N_G(X') \setminus S$.

Using $C(X'; S) = 0$ together with the definition of C, the observation above, and (2) which implies $v \notin X'$; we find

$$C(X''; S) = C(X''; S) - C(X'; S) = -d + \deg_S(v) < 0.$$

However, we have $|X''| \leq m + 1 \leq 2m$, so $C(X''; S) \geq 0$ since S is (d, m)-good, a contradiction to the inequality above. We conclude the result. $\square$

27.2 Variants of the Method

27.2.1 Sharpening Friedman-Pippenger

While the proof above is nice and clean, there is some slack that can be tightened to give stronger conclusions. For example, a slight aesthetically displeasing feature of Theorem 27.1 is that it requires our graph G to be $(d + 1, 2m)$-expanding in order to find trees of maximum degree d, and it heuristically feels like one should be able to reduce the $d + 1$ down to just d since, for example, this is all that is needed to find stars $K_{1,d}$.

By reexamining the proof above, we see that the only place we used the $d + 1$ property was in Claim 27.4 to show that critical sets always have size at most m. However, a quick inspection shows that we don't need the full power of $(d + 1, 2m)$-expanding for this proof to go through; all we really need is something slightly stronger than $(d, 2m)$-expansion for sets of size at least $m + 1$. With this observation in mind, an essentially identical proof can be used to prove the following strengthened version of the Friedman-Pippenger Theorem.

Theorem 27.6 ([13],[86])**.** *Let d, m, M be positive integers and G a graph such that the following holds:*

- *Every $X \subseteq V(G)$ with $0 < |X| \leq m$ has $|N_G(X)| \geq d|X| + 1$,*
- *Every $X \subseteq V(G)$ with $m < |X| \leq 2m$ has $|X_G(X)| \geq d|X| + M$.*

Then G contains every tree with M vertices.

Theorem 27.6 was first explicitly stated by Balogh, Csaba, Pei, and Samotij [13] who realized that it was an easy consequence of a substantial extension of Theorem 27.1 due to Haxell [86] which will be discussed in the following subsection.

Observe that G being $(d+1, 2m)$-expanding implies it satisfies the hypothesis of Theorem 27.6 with $M = m+1$, so Theorem 27.6 does indeed recover Theorem 27.1. In fact, it is quite a bit stronger: observe that with Theorem 27.1 alone, one can't hope to embed trees of size larger than $|V(G)|/2(d+1)$ (since sets of size more than $|V(G)|/(d+1)$ can't expand by a factor of at least $(d+1)$). In contrast, Theorem 27.6 can be used to embed trees of size $(1-o(1))|V(G)|$ provided G has sufficiently strong expansion properties.

Theorem 27.7. *Let $\mathcal{T}_{d,N}$ denote the set of trees of maximum degree d on at most N vertices. For all $d, \epsilon > 0$, there exists some $C > 0$ such that $G_{n,p}$ with $p = C/n$ contains every member of $\mathcal{T}_{d,(1-\epsilon)n}$ with high probability.*

This result was originally proven by Alon, Krieleveich, and Sudakov [7] through a direct application of the Friedman-Pippenger Theorem. This result was reproven by Balogh, Csaba, Pei, and Samotij [13] using Theorem 27.6 to give a shorter proof with a better dependency on C. We present a short argument in the spirit of [13] that can be found in notes of Liu (which also discuss many other techniques for finding structures in pseudo-random graphs).

Before we get into the proof, we observe that there is no hope in applying Friedman-Pippenger or any of its variants to $G_{n,p}$ with $p = C/n$. Indeed, in this range $G_{n,p}$ will contain a linear number of isolated vertices in expectation, and any subset of these vertices will fail to have the necessary expansion properties. Given this, our only hope of using Friedman-Pippenger type results is to restrict to an induced subgraph of $G_{n,p}$ after removing a linear set of vertices which have bad expansion properties. And indeed, the following says that if a graph G has a given amount of expansion for sets of a given size m, then one can delete a small set from G to achieve a comparable level of expansion for all sets of size at most m. For technical reasons this result will require us to work with a subset of $N(X)$ instead of $N(X)$ itself.

Lemma 27.8. *Let G be a graph and define $N'_G(X) := N_G(X) \setminus X$ for all $X \subseteq V(G)$. If $D > 0$ and m is an integer such that every $X \subseteq V(G)$ of size m satisfies $|N'_G(X)| \geq D|X|$, then there exists a set $B \subseteq V(G)$ of size less than m such that $|N'_{G-B}(X)| \geq \left(\frac{D}{2} - 1\right)|X|$ for all $X \subseteq V(G-B)$ of size at most m.*

Proof. Let B be a maximal set of size at most $2m$ with the property that $|N'_G(B)| < \left(\frac{D}{2} - 1\right)|B|$. We claim that $|B| < m$. Otherwise, there would exist some $B' \subseteq B$ of size exactly m, and the hypothesis of the lemma implies

$$|N'_G(B)| \geq |N'_G(B')| - |B| \geq M|B'| - |B| \geq \frac{D}{2}|B| - |B| = \left(\frac{D}{2} - 1\right)|B|,$$

where this last inequality used $|B| \leq 2m = 2|B'|$. This gives a contradiction to how we defined B.

Now assume for contradiction that there exists some $X \subseteq V(G-B)$of size at most m with $|N'_{G-B}(X)| < \left(\frac{D}{2} - 1\right)|X|$. In this case we observe that $X \cup B$ is a set of size at most $2m$ such that

$$|N'_G(X \cup B)| \leq |N'_{G-B}(X)| + |N'_G(B)| < \left(\frac{D}{2} - 1\right)|X| + \left(\frac{D}{2} - 1\right)|B| = \left(\frac{D}{2} - 1\right)|X \cup B|.$$

□

We now use this to prove random graphs contain nearly spanning trees.

Proof. Theorem 27.7 Let C be sufficiently large in terms of ϵ, d and take $m := \frac{100\log(1/\epsilon)}{C}n$. It is striaghtforward to show using the Chernoff bound that with high probability we have $|N'_{G_{n,p}}(X)| \geq (1-\epsilon/2)n$ for all X of size exactly m. By applying Lemma 27.8 with $D = (1-\epsilon/2)n/m \geq 2d+3$ (this inequality using C sufficiently large in terms of ϵ, d), there exists some $B \subseteq V(G_{n,p})$ such that after deleting B, we have for all sets X of size at most m that

$$|N_{G_{n,p}-B}(X)| \geq |N'_{G_{n,p}-B}(X)| \geq (d+1)|X| \geq d|X| + 1.$$

On the other hand, if X has size between $m+1$ and $2m$ then we can take $X' \subseteq X$ to be a subset of size exactly m and conclude

$$|N_{G_{n,p}-B}(X)| \geq |N'_{G_{n,p}}(X')| - |B| \geq (1-\epsilon/2)n - m = (\frac{\epsilon}{2}n - m) + (1-\epsilon)n \geq d|X| + (1-\epsilon)n,$$

where this last step used $|X| \leq 2m$ and that C is sufficiently large in terms of ϵ, d. By Theorem 27.6 we conclude that $G_{n,p} - B$ contains a copy of every $T \in \mathcal{T}_{d,(1-\epsilon)n}$, proving the result. □

27.2.2 Haxell's Method

As mentioned above, Theorem 27.6 is a corollary of a more involved result of Haxell's. We will not state her result in full here, but we will briefly discuss her key innovation of building trees T through a series of intermediate "stages." The rough idea is that we begin by building $S = K_1$, then we go and build $S = T_1$ for some tree T_1 obtained from K_1 by adding leaves, then we build $S = T_2$ in a similar way and so on. Crucially, if some tree T_i for $i \geq 1$ contains every vertex of T which has degree d, then from this point onward no vertex we add on to S will ever need more than $d-1$ leaves added to it. As such, we can replace the $-d|X|$ term in the definition of $C(X; S)$ by something like $-(d-1)|X|$, making it easier to find S for which $C(X; S)$ is sufficiently large.

As far as we are aware, Haxell's full technical result is not used much in the literature, with most papers either using the simplified Theorem 27.6 or the extendability method discussed below. That being said, we note that her full result is the only version of Friedman and Pippenger's approach that we are aware of which gives the following nice variant of the Erdős-Sós conjecture for hosts which avoid $K_{2,r}$'s.

Theorem 27.9 ([86])**.** *If t is a positive integer and $r = \lceil t/18 \rceil$, then every graph with average degree greater than $t-1$ which is $K_{2,r}$-free contains every tree with t edges as a subgraph.*

27.2.3 The Extendability Method: Intuition

Perhaps the most used variant of the Friedman-Pippenger Theorem is an approach known as the extendability method. We begin with an entirely optional subsection discussing the intuition behind how one might come across this approach given the proof of Friedman and Pippenger, after which we have a more formal subsection discussing of the relevant definitions and lemmas of the method together with a very brief survey of what this method can be used for.

With an eye towards trying to sharpen the proof of Theorem 27.1, one might come to the realization that the same proof goes through if we replace the function

$$C(X;S) := |N_G(X) \setminus S| - d|X| + \sum_{x \in X \cap V(S)} \deg_S(x)$$

with the function

$$\begin{aligned} D(X;S) &:= |N_G(X) \setminus S| - (d-1)|X| - |X \cap S| + \sum_{x \in X \cap V(S)} \deg_S(x) \\ &= |N_G(X) \setminus S| - (d-1)|X| + \sum_{x \in X \cap V(S)} (\deg_S(x) - 1). \end{aligned}$$

As to why exactly you might come up with this, we recall in the proof of Theorem 27.1 that we obtained the structural fact (3) of having $\ell \notin X_\ell$ for all $\ell \in L$, which followed from the fact that $C(X;S)$ increases if we add in a new term with $\deg(\ell) = 1$ to its sum. That being said, we never actually used this structural fact in our proof.

As such, the same proof would almost go through if we subtracted from $C(X;S)$ the additional term $|X \cap S|$ (since doing so would only change that we can no longer guarantee (3), but again this was not used in the proof). However, a careful inspection shows that this change would make the proof of Claim 27.4 fail to go through in the very particular case where $S \subseteq X$ and $|S| = m$. This problem in turn can be patched by replacing the $-d|X|$ term in $C(X;S)$ with $-(d-1)|X|$. Note crucially that the only place we used the exact coefficient in front of the $-|X|$ term was at the very end when we computed $C(X'';S) - C(X';S)$. However, because we are also now subtracting the term $|X \cap S|$ in our function D, this difference will be $(d-1) - (\deg_S(v) - 1) = d - \deg_S(v)$ exactly as before, giving the result as before.

Offhand there doesn't appear to be any advantage in working with $D(X;S)$ over our original $C(X;S)$. However, its exact formulation (and in particular the subtraction of $|X \cap S|$) will be crucial in several of the key lemmas of the extendability method discussed below.

27.2.4 The Extendability Method: Formal Definitions

Motivated by our discussion above, the key idea will be to change our definition of "good" subgraphs to be with respect to a new cost function $D(X;S)$ which was originally introduced by Glebov, Johannsen and Krivelevich [78].

Definition 4. For positive integers d, m with $d \geq 3$, we say that a subgraph S of a graph G is (d,m)-extendable if S has maximum degree at most d and if for every set $X \subseteq V(G)$ with

$|X| \leq 2m$ we have

$$D(X;S) := |N_G(X) \setminus S| - (d-1)|X| + \sum_{x \in X \cap S} (\deg_S(x) - 1) \geq 0.$$

It's not clear to me why $d \geq 3$ is used here, but that's how it's typically defined.

In order to use the notion of extendability, we will need lemmas analogous to Lemma 27.2 saying that we can build new extendable subgraphs from old ones. We highlight a few such lemmas here whose names come from [78]. We begin with an analog of Lemma 27.2 whose proof is nearly word for word the same.

Lemma 27.10 (Vertex Extension Lemma)**.** *Let d, m be positive integers with $d \geq 3$ and let S be a (d,m)-extendable subgraph of G. If every subset $X \subseteq V(G)$ with $m \leq |X| \leq 2m$ has*

$$|N_G(X)| \geq |S| + 2dm + 1,$$

then for every $v \in S$ with $d_S(v) \leq d-1$ there exists $\ell \in N_G(s) \setminus S$ such that $S + v\ell$ is (d,m)-extendable.

The previous lemma allows us to add leaves to extendable subgraphs. The next lemma allows us to "rollback" this process, which (despite perhaps looking inconsequential at first glance) is to a large extent the main strengthening of the extendability method over Friedman and Pippenger's original approach.

Lemma 27.11 (Removal Lemma)**.** *Let d, m be positive integers with $d \geq 3$ and let S be a subgraph of G. If there exist vertices $v \in S$ and $\ell \in N_G(S) \setminus S$ adjacent in G such that the subgraph $S + v\ell$ is (d,m)-extendable, then S is (d,m)-extendable as well.*

Here we emphasize that because $\ell \notin S$ that ℓ is necessarily a leaf of $S + v\ell$.

Proof. For any X we observe that

$$D(X;S) - D(X;S+v\ell) = \mathbb{1}[\ell \in N_G(X)] - \mathbb{1}[v \in X].$$

Since $v\ell \in E(G)$ by hypothesis, having $v \in X$ implies $\ell \in N_G(X)$, and hence this difference is always non-negative, showing $D(X;S) \geq 0$ whenever we have $D(X;S+v\ell) \geq 0$, proving the result. $\square$

We emphasize that the argument above would not work if we were using $C(X;S)$ instead of $D(X;S)$, as in this case the difference would have an additional $-\mathbb{1}[\ell \in X]$ term which we wouldn't be able to control.

We next consider two lemmas that allow us to create extendable structures which contain cycles.

Lemma 27.12 (Edge Insertion Lemma)**.** *Let d, m be positive integers with $d \geq 3$ and let S be a (d,m)-extendable subgraph of G. If $u, v \in S$ are such that $\deg_S(u), \deg_S(v) \leq d-1$ and $uv \in E(G)$, then $S + uv$ is a (d,m)-extendable subgraph of G.*

Indeed, this just follows from observing that $D(X;S)$ can never decease if we add an edge between two vertices of S.

Lemma 27.13 (Weak Connection Lemma)**.** *Let d, m be positive integers with $d \geq 3$ and let G be an n-vertex graph such that $e_G(X,Y) \geq 2$ for all disjoint $X, y \subseteq V(G)$ with $|X|, |Y| \geq m$ and let S be a (d,m)-extendable subgraph on at most $n - 10dm$ vertices.*

Let $k = \lceil \log(2m)/\log(d-1) \rceil$. Let a, b be two vertices in S which have degree at most $d/2$ in S. Then there exists a path P of length $2k+1$ in G with endpoints a and b such that all vertices of P except the two endpoints do not belong to S and such that $S+P$ is (d,m)-extendable.

The actual Connection Lemma used in practice further allows one to replace the vertices a, b with sets of vertices A, B and then guarantees that we can find a path between some vertex of A to some vertex of B, but for ease of presentation we only consider this weaker version.

Sketch of Proof. By using the Vertex Extension lemma, we can attach to a $\lceil (d-1)/2 \rceil$ edges ax_i and then build from each x_i a $(d-1)$-ary tree of length $k-1$ while maintaining that our subgraph is extendable (we can do this both because we assume $\deg_S(a) \leq d/2$ and because we assume G has a suitable expansion property).

We perform the exact same procedure on b. Since the total number of leaves for the trees attached to each of a, b is at least m by definition of k, the expansion property of G guarantees there is an edge between two leaves, and we can add this and maintain extendability due to the Edge Insertion Lemma. Finally, the Removal Lemma allows us to strip away the leaves until we are only left with $S+P$ as desired. $\square$

27.3 Applications

While the extendability method and its variants are used in a number of important applications, most of these applications require quite a bit of technical analysis. Because of this, we will only briefly survey some of the results that are possible with these methods and refer the interested reader to the individual papers for more details on how the methods are used.

Historically, the original motivation of the Friedman-Pippenger theorem was for a problem in size Ramsey theory. To motivate this area, given two graphs F, G, we write $G \to F$ if every 2-coloring of the edges of G contains a monochromatic copy of F, and we then define the Ramsey number

$$r(F) := \min\{|V(G)| : G \to F\}.$$

Note that we could have equivalently defined $r(F)$ to be the smallest n such that $K_n \to F$, but this definition allowing for more general hosts other than just K_n motivates interesting generalizations. In particular, we define the *size-Ramsey number*

$$\hat{r}(F) := \min\{|E(G)| : G \to F\}.$$

Note that we trivially have $\hat{r}(F) \leq \binom{r(F)}{2}$ by taking G to be a complete graph on $r(F)$ vertices. A remarkable argument by Chvátal [57] shows that this bound is tight for cliques $F = K_r$ (despite the fact that we have no idea what $r(K_r)$ is!). However, for other F there can be a substantial gap between $\hat{r}(F)$ and $\binom{r(F)}{2}$. Notably, Friedman and Pippenger proved the following, extending a previous result of Beck for paths [18].

Theorem 27.14 ([70])**.** *If F is an n-vertex tree with maximum degree d, then $\hat{r}(F) = O_d(n)$.*

Proof Sketch. For this, it suffices to show that there exists a graph G with $O_d(n)$ edges such that any 2-coloring of its edges contains a monochromatic copy of T. In fact we will prove the stronger fact that any subgraph $G' \subseteq G$ containing at least half of the edges of G contains a copy of T.

With the above goal in mind, we let G be a "sparse optimally pseudorandom graph", i.e. a regular graph of bounded degree whose eigenvalues besides λ_1 are all small. Such graphs are known to exist and have effective expansion properties. Friedman and Pippenger further prove that if you restrict to some subgraph $G' \subseteq G$ which contains only, say, half the edges of G, then one can find an induced subgraph of G' which is $(d+1, 2n)$-expanding and hence contains a copy of T by Theorem 27.1, proving the result. $\square$

This result was further refined by Dellamonica [48] who (together with a lower bound from Beck [18]) showed that $\hat{r}(T) = \Theta(\beta(T))$ for every tree T, where $\beta(T)$ is some slightly complicated invariant introduced by Beck.

The construction of Dellamonica is rather simple: he just takes G to be a p-random subgraph of an unbalanced complete bipartite graph and then argues that with high probability any constant proportion of the edges of G will contain a copy of T. There are two main improvements of his analysis over that of Friedman and Pippenger. First, he tailors his argument specifically to random-like graphs (rather than just graphs satisfying some relatively weak expansion property). Second, he adds algorithmic component to the proof by (roughly speaking) taking each vertex x with relatively few neighbors and constructing some subset $R_x \subseteq V(G)$ of "reserved" vertices which may only be used as neighbors of x in the embedding (if x gets chosen). However, the full details of his argument are quite involved.

In addition to size-Ramsey numbers, another major area of application for Friedman-Pippenger type theorems come from finding structures is random and pseudo-random graphs. We already saw one example in Theorem 27.7 where we gave an easy proof showing that $G_{n,p}$ with $p = C/n$ contains every bounded-degree tree on $(1-\epsilon)n$ vertices with high probability. With substantially more work, Montgomery [118] managed to show that $G_{n,p}$ at $p = C \log n/n$ contains every bounded-degree *spanning* tree with high probability. While the paper is very long and involves many ideas, one of the central tools of the argument is in using the extendability method, and in particular a variant of the Connection Lemma. It is also worth noting that, although the original proof of Montgomery is quite difficult, his result is now essentially a one page consequence of the more modern Spreadness Theorem Theorem 8.1.

Possibly include a longer list of references for people to look at for examples.

28 The Delta-System Method

Recall that a *k-sunflower* (also called a delta-system) is a hypergraph $\mathcal{S}$ with edges $e_1, \ldots, e_k$ such that there exists a set K called the *kernel* which has $e_i \cap e_j = K$ for all $i \neq j$. Roughly speaking, the Delta-system method is any proof using the following observation, which is usually credited to Deza, Erdős, and Frankl [49].

Lemma 28.1. *If $\mathcal{H}$ is an r-graph which contains an $(r+1)$-sunflower with kernel K, then for every edge $e \in \mathcal{H}$, there exists an edge $f \in \mathcal{H}$ with $e \cap f \subseteq K$.*

Proof. Let $e_1, \ldots, e_{r+1}$ be the edges of the sunflower. Since each of the sets $e_1 \setminus K, \ldots, e_{r+1} \setminus K$ are non-empty disjoint sets, one of these $e_i \setminus K$ sets must be disjoint from e. Taking $f = e_i$ gives the result. □

An effective tool to use in conjunction with this observation is Füredi's intersection semilattice lemma (which itself is proven using the Delta-system method). The full statement is a little intimidating, so we'll start by just stating a consequence of it.

Lemma 28.2 (Weak intersection semilattice lemma)**.** *For all r, s, there exists $c = c(r,s) > 0$ such that for every r-graph $\mathcal{H}$, there exists a subgraph $\mathcal{H}' \subseteq \mathcal{H}$ with $|\mathcal{H}'| \geq c|\mathcal{H}|$ such that for all $e, f \in \mathcal{H}'$, $e \cap f$ is the kernel of an s-sunflower.*

That is, we can approximate $\mathcal{H}$ by a hypergraph $\mathcal{H}'$ such that any two edges of $\mathcal{H}'$ are petals of a large sunflower. This quickly gives the following strengthening of the Erdős-Rado sunflower lemma due to Mubayi and Zhao [121].

Corollary 28.3 ([121])**.** *For all r, s there exists a constant $C = C(r,s)$ such that if $\mathcal{H}$ is an n-vertex r-graph with $|\mathcal{H}| \geq Cn^{r-t-1}$, then $\mathcal{H}$ contains an s-sunflower which has core of size at most t.*

Proof. We may assume n is sufficiently large in terms of r, as otherwise one can trivially find a sufficiently large C. Let $C = 2c^{-1}$ with c the constant from the previous lemma. Then $|\mathcal{H}'| \geq 2n^{r-t-1} > \binom{n-t-1}{r-t-1}$. By the Erdős-Ko-Rado theorem for $(t+1)$-intersecting hypergraphs (see Theorem 10.4), $\mathcal{H}'$ must contain two edges e, f which intersect in less than $t+1$ vertices. By assumption $e \cap f$ is the core of a sunflower with at least s petals, proving the result. □

We'll now state the full intersection semilattice lemma. For this, if $\mathcal{H}$ is an r-partite r-graph with partition $\bigcup V_i$ and if S is a set of vertices, then we define $\mathrm{proj}(S) := \{i : S \cap V_i \neq \emptyset\}$. That is, $\mathrm{proj}(S)$ records which coordinates its vertices are in. Given a hypergraph $\mathcal{H}$ and a set of vertices S, define $d^*(S)$ to be the largest integer d such that there exist edges $e_1, \ldots, e_d \in \mathcal{H}$ with $e_i \cap e_j = S$ for all $i \neq j$. In other words, $d^*(K)$ is the size of the largest sunflower which contains K as its kernel.

It's unfortunate that here $\mathcal{H}'$ is the part we can approximate while it's the opposite for spread approximations. Probably change one of these, most likely the spread approximation one.

Lemma 28.4 (Intersection semilattice lemma)**.** *For all r, s, there exists $c = c(r, s) > 0$ such that for every r-graph $\mathcal{H}$, there exists an r-partite subgraph $\mathcal{H}' \subseteq H$ with $|\mathcal{H}'| \geq c|\mathcal{H}|$ and a hypergraph $\mathcal{J} \subseteq 2^{[r]}$ not containing the edge of size r such that:*

(1) $\mathcal{J}$ is intersection closed, i.e. $I, J \in \mathcal{J}$ implies $I \cap J \in \mathcal{J}$.

(2) For every $e \in \mathcal{H}'$, $\{\mathrm{proj}(e \cap f) : f \in \mathcal{H}' \setminus \{e\}\} = \mathcal{J}$.

(3) $d^*_{\mathcal{H}'}(e \cap f) \geq s$ for all $e, f \in \mathcal{H}'$.

Note that if we ignore (1) and (2) we get back Lemma 28.2. Roughly speaking, this result says that we can approximate a large chunk of $\mathcal{H}$, namely $\mathcal{H}'$, by a small hypergraph $\mathcal{J}$ such that for any edge $e \in \mathcal{H}'$, the hypergraph $\mathcal{J}$ tells you exactly how other edges can intersect e, and moreover, (3) guarantees that each possible intersection occurs at least s times. We say that an $\mathcal{H}'$ as in the conclusion of this lemma is $(s, \mathcal{J})$-homogeneous.

We postpone proving this result for the moment and instead look at some consequences. For $\mathcal{J} \subseteq 2^{[r]} \setminus [r]$, define the rank

$$\mathrm{rank}(\mathcal{J}) = \min\{|T| : T \subseteq [r],\ T \not\subseteq I\ \forall I \in \mathcal{J}\},$$

i.e. this is the smallest integer t such that there exists a t-set not contained in an edge of $\mathcal{J}$. For example, $\mathrm{rank}(\mathcal{J}) > 1$ if and only if every vertex of $[r]$ is contained in an edge of $\mathcal{J}$.

Lemma 28.5. *If $\mathcal{H}'$ is an n-vertex $(s, \mathcal{J})$-homogeneous r-graph, then $|\mathcal{H}'| \leq \binom{n}{\mathrm{rank}(\mathcal{J})}$.*

Note that s does not appear in this bound. Before looking at the proof, the reader may want to try proving this result for themselves when $\mathrm{rank}(\mathcal{J}) = 1$ in order to get a sense for the definitions.

Proof. Let $T \subseteq [r]$ be a set such that $|T| = \mathrm{rank}(\mathcal{J})$ and such that $T \not\subseteq J$ for all $J \in \mathcal{J}$. Given an edge $e \in \mathcal{H}'$, let $\phi(e) = e \cap \bigcup_{i \in T} V_i$. We claim that ϕ is injective. Indeed, if $\phi(e) = \phi(f)$, then $T \subseteq \mathrm{proj}(e \cap f) \in \mathcal{J}$, a contradiction to our assumption on T. Since ϕ maps edges of $\mathcal{H}'$ injectively to sets of size $\mathrm{rank}(\mathcal{J})$, we conclude the result. □

We can use this result to give yet another proof of the Erdős-Ko-Rado theorem for t-intersecting hypergraphs, whose statement we recall below.

Theorem 28.6. *Let $\mathcal{H}$ be an n-vertex r-graph such that $|e \cap f| \geq t$ for all $e, f \in \mathcal{H}$. If n is sufficiently large in terms of r, then $|\mathcal{H}| \leq \binom{n-t}{r-t}$ with equality holding if and only if $\mathcal{H}$ consists of every edge containing some fixed set T of size t.*

Proof. Apply Lemma 28.4 with $s = r + 1$ and let $\mathcal{H}', \mathcal{J}$ be the resulting hypergraphs with $\bigcup V_i$ the r-partition of $\mathcal{H}'$. First note that if $\mathrm{rank}(\mathcal{J}) < r - t$, then by Lemma 28.5 we have

$$|\mathcal{H}| \leq c^{-1}|\mathcal{H}'| \leq c^{-1}\binom{n}{r-t-1} < \binom{n-t}{r-t},$$

where this last step holds for n sufficiently large in terms of $c = c(r, r+1)$. Thus if $|\mathcal{H}| \geq \binom{n-t}{r-t}$, we must have $\mathrm{rank}(\mathcal{J}) \geq r - t$.

Let S be with $|S| = \mathrm{rank}(\mathcal{J})$ such that no edge of $\mathcal{J}$ contains S, and let $T = [r] \setminus S$. By the above we may assume $|S| \geq r - t$, and hence $|T| \leq t$. By definition of $\mathrm{rank}(\mathcal{J}) = |S|$, for every $i \in S$, there exists an edge $J_i \in \mathcal{J}$ such that $S \setminus \{i\} \subseteq J_i$. Note that $i \notin J_i$ by assumption of S not being contained in any edge of $\mathcal{J}$, which means $J := \bigcap_{i \in S} J_i \subseteq T$. Because $\mathcal{J}$ is intersection closed, we have $J \in \mathcal{J}$.

We claim that $|J| \geq t$. Indeed, by definition of $J \in \mathcal{J}$, there exist two edges of $\mathcal{H}' \subseteq \mathcal{H}$ whose intersection is exactly J, and by the t-intersecting property we must have $|J| \geq t$. Because $|T| \leq t$ and $J \subseteq T$, we conclude that $J = T \in \mathcal{J}$.

Now let $e \in \mathcal{H}'$ and $K = e \cap \bigcup_{i \in T} V_i$, noting that $|K| = t$. Then Lemma 28.4 guarantees that there is a sunflower with at least $r + 1$ petals and K as its kernel. This implies that for every edge $f \in \mathcal{H}$, there exists an edge $e' \in \mathcal{H}'$ which contains K and which is disjoint from $f \setminus K$. Thus to have $|e' \cap f| \geq t$, we must have $K \subseteq f$. In other words, every edge of $\mathcal{H}$ must contain the t-set K. This implies the result. □

The above argument actually gives the following stability result: for all r, t there exists a constant $c' = c'(r, t)$ such that if $\mathcal{H}$ is t-intersecting with $|\mathcal{H}| > c'\binom{n-t}{r-t}$, then there exists a set of size t which is contained in every edge of $\mathcal{H}$.

Remark 28.7. Stronger versions of Theorem 10.4 are known. For example, Wilson [146] determined the maximum size of a t-intersecting family for *any* value n. In another direction, Frankl and Füredi [65] showed that the conclusion of the theorem holds if we only impose the hypothesis $|e \cap f| \neq t$ for $e \neq f$ provided $r \geq 2t + 2$, with their proof using a somewhat more involved version of the Delta-system method.

A similar argument works for more general kinds of intersection problems. Given a set $L \subseteq \{0, 1, \ldots, r-1\}$, we say that a hypergraph $\mathcal{H}$ is an (n, r, L)-system if it's an n-vertex r-graph such that $|e \cap f| \in L$ for all $e, f \in \mathcal{H}$ distinct. For example, $(n, r, \{t, t+1, \ldots, r\})$-systems are t-intersecting hypergraphs. Little is known about how large (n, r, L)-systems can be for general L, but one can get effective bounds in terms of ranks. To this end, for any $L \subseteq \{0, 1, \ldots, r-1\}$, define

$$\mathrm{rank}(r, L) = \max_{\mathcal{J}} \mathrm{rank}(\mathcal{J}),$$

where the maximum ranges over all $\mathcal{J} \subseteq 2^{[r]}$ containing no edges of size r which are intersection-closed with $|J| \in L$ for all $J \in \mathcal{J}$.

Theorem 28.8. *If $\mathcal{H}$ is an (n, r, L)-system, then*

$$|\mathcal{H}| = O(n^{\mathrm{rank}(r,L)}).$$

Proof. Let $\mathcal{H}', \mathcal{J}$ be as in Lemma 28.4. Observe that $\mathcal{J}$ is intersection closed and that $|J| \in L$ for all $J \in \mathcal{J}$ (since otherwise two edges of $\mathcal{H}' \subseteq \mathcal{H}$ would fail to have $|e \cap e'| \in L$). Thus letting c be the constant from Lemma 28.4, we have

$$|\mathcal{H}| \leq c^{-1} e(H') \leq c^{-1}\binom{n}{\mathrm{rank}(\mathcal{J})} \leq c^{-1}\binom{n}{\mathrm{rank}(r, L)},$$

where this second inequality used Lemma 28.5 and the last inequality used the definition of $\mathrm{rank}(r, L)$. We conclude the result. $\square$

It's conjectured by Frankl that for all r, L there exist (n, r, L)-systems of size $\omega(n^{\mathrm{rank}(r,L)-1})$. This is unknown in general, but see e.g. [66, Theorem 16.6] for a construction of size $\Omega(n^{1+1/(r-1)})$ whenever $\mathrm{rank}(r, L) \geq 2$.

One of the best general bounds on the size of (n, r, L)-systems is the Deza-Erdős-Frankl theorem, which states that such a system $\mathcal{H}$ satisfies

$$|\mathcal{H}| \leq \prod_{\ell \in L} \frac{n-\ell}{r-\ell} = O(n^{|L|}),$$

provided n is sufficiently large. One can prove this result using a variant of the Delta-system method, though we omit doing so here. Instead, we give an easy proof of the asymptotic result by utilizing the following.

Lemma 28.9. *Every* $L \subseteq \{0, 1, \ldots, r-1\}$ *satisfies* $\mathrm{rank}(r, L) \leq |L|$.

Proof. We prove the result by induction on $|L|$, the case $|L| = 0$ being trivial. Assume we have proven the result for all L with $|L| < k$. We first consider the case $|L| = k$ and $0 \in L$. Let $\mathcal{J}$ be an intersection closed hypergraph on $2^{[r]} \setminus [r]$ with $|J| \in L$ for all $J \in \mathcal{J}$ and $\mathrm{rank}(\mathcal{J}) = \mathrm{rank}(r, L)$. For any vertex $x \in [r]$, let $\mathcal{J}_x = \{J - x : x \in J \in \mathcal{J}\}$ be the link hypergraph. If $L' = \{\ell - 1 : \ell \in L,\ \ell \neq 0\}$, then we see that $\mathcal{J}_x$ is intersection closed with $|J| \in L'$ for all $J \in \mathcal{J}_x$. Because $|L'| = |L| - 1$, our inductive hypothesis implies that $\mathrm{rank}(r-1, L') \leq |L| - 1$, i.e. there exists some set T of size $|L| - 1$ in $[r] \setminus \{x\}$ which is not contained in any edge of $\mathcal{J}_x$, which implies $T \cup \{x\}$ is a set of size $|L|$ not contained in any edge of $\mathcal{J}$. We conclude $\mathrm{rank}(r, L) = \mathrm{rank}(\mathcal{J}) \leq |L|$.

Now assume $0 \notin L$, and again let $\mathcal{J}$ be an intersection closed hypergraph on $2^{[r]} \setminus [r]$ with $|J| \in L$ for all $J \in \mathcal{J}$ with $\mathrm{rank}(\mathcal{J}) = \mathrm{rank}(r, L)$. Let $I = \bigcap_{J \in \mathcal{J}} J$. Note that by definition I is contained in every edge of $\mathcal{J}$, and by the intersection closed property we have $I \in \mathcal{J}$, and hence $|I| \in L$. Define $L' = \{\ell - |I| : \ell \geq |I|\}$, and note that the link hypergraph $\mathcal{J}_I = \{J \setminus I : J \in \mathcal{J}\}$ has all of its edge sizes lying in L'. Since $|L'| \leq |L|$ and $0 \in L'$, the previous case implies $\mathrm{rank}(\mathcal{J}_I) \leq |L|$, which implies there exists some set $J \subseteq [r] \setminus I$ of size L not contained in an edge of $\mathcal{J}_I$, and hence this set continues to not be contained in an edge of $\mathcal{J}$ (since every edge of $\mathcal{J}$ is the union of an edge of $\mathcal{J}_I$ with I). We conclude $\mathrm{rank}(r, L) = \mathrm{rank}(\mathcal{J}) \leq |L|$, proving the result. $\square$

This together with the previous theorem immediately gives the following.

Corollary 28.10. *If* $\mathcal{H}$ *is an* (n, r, L)*-system, then*

$$e(H) = O(n^{|L|}).$$

It is not difficult to show that this result is tight if $L = \{0, 1, \ldots, t-1\}$.

Before moving on, we note that while all of our applications here came from extremal set theory, the intersection semilattice lemma has application to other areas of extremal combinatorics as well. See for example [120], where this lemma is used to bound the Turán number of a class of linear hypergraphs called "expansions."

28.1 Proof of the Semilattice Intersection Lemma

We first recall that for every r-graph $\mathcal{H}$ there exists a subgraph $\mathcal{H}' \subseteq \mathcal{H}$ which is r-partite and which keeps a constant proportion of its edges. Thus it suffices to prove that if $\mathcal{H}$ has r-partition $\bigcup V_i$, then one can find a subgraph $\mathcal{H}' \subseteq \mathcal{H}$ and $\mathcal{J} \subseteq 2^{[r]} \setminus [r]$ such that $|\mathcal{H}'| \geq c(r,s)|\mathcal{H}|$ and

(1) $\mathcal{J}$ is intersection closed, i.e. $I, J \in \mathcal{J}$ implies $I \cap J \in \mathcal{J}$.

(2) For every $e \in \mathcal{H}'$, $\{\mathrm{proj}(e \cap f) : f \in \mathcal{H}' \setminus \{e\}\} = \mathcal{J}$.

(3) $d^*_{\mathcal{H}'}(e \cap f) \geq s$ for all $e, f \in \mathcal{H}'$.

Claim 28.11. *If there exists $\mathcal{H}', \mathcal{J}$ satisfying (2) and (3) with $s \geq r+1$, then they automatically satisfy (1).*

Proof. Let $J_1, J_2 \in \mathcal{J}$. This means that for any edge $e \in \mathcal{H}'$, there exist edges e_1, e_2 such that e, e_i intersect exactly in the coordinates of J_i and that this intersection is the kernel of a sunflower in $\mathcal{H}'$ on at least $r+1$ petals. In particular, there must exist an edge $f \in \mathcal{H}'$ which contains $e \cap e_1$ and which is otherwise disjoint from e_2 (namely, f is one of the edges of the sunflower with core $e \cap e_1$). With this $\mathrm{proj}(f \cap e_2) = J_1 \cap J_2$, so necessarily $J_1 \cap J_2 \in \mathcal{J}$. □

With this claim in mind, we only have to find $\mathcal{H}', \mathcal{J}$ satisfying (2) and (3) (this is immediate if $s \geq r+1$, and for all other values of s we can take $c(r,s) = c(r, r+1)$ and apply the $s = r+1$ result). For the rest of the proof, given a hypergraph $\mathcal{H}$, we define

$$\mathcal{I}(\mathcal{H}) = \{\mathrm{proj}(e \cap f) : e, f \in \mathcal{H},\ e \neq f\}.$$

Note that if (2) are (3) are satisfied for some $\mathcal{J}$, then it must be that $\mathcal{J} = \mathcal{I}(\mathcal{H})$.

Claim 28.12. *For any r-partite r-graph $\mathcal{H}$, one can decompose $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}_0 \cup \bigcup_{I \in \mathcal{I}(\mathcal{H})} \mathcal{H}_I$ such that $\mathcal{H}_0$ satisfies (2) and (3) for some $\mathcal{J}$, and such that $I \neq \mathrm{proj}(K)$ for any set K which is the kernel of a sunflower on at least s petals in $\mathcal{H}_I$.*

Proof. Initially start with $\mathcal{H}_0 = \mathcal{H}$ and $\mathcal{H}_I = \emptyset$ for all $I \in \mathcal{I}$. Consider the following procedure. If at any point $\mathcal{H}_0$ satisfies (2) and (3) for some set $\mathcal{J}$, then we stop and output the current sets. Otherwise, it is not difficult to see that there must exist some edge $e \in \mathcal{H}_0$ and $K \subseteq e$ such that $\mathrm{proj}(K) \in \{\mathrm{proj}(f \cap g) : f, g \in \mathcal{H}_0,\ f \neq g\}$ but K is not the kernel of a sunflower with at least s petals in $\mathcal{H}_0$ (otherwise the conditions would be satisfied with $\mathcal{J} = \{\mathrm{proj}(f \cap g) : f, g \in \mathcal{H}_0,\ f \neq g\}$). Delete e from $\mathcal{H}_0$ and add it to $\mathcal{H}_{\mathrm{proj}(K)}$.

We claim that this procedure gives the desired result. Indeed, $\mathcal{H}_0$ satisfies (2) and (3) by construction. Assume for contradiction that there existed some I with $\mathcal{H}_I$ containing a sunflower on at least s petals with kernel K satisfying $\mathrm{proj}(K) = I$. Let e be the first edge of this sunflower that was added to $\mathcal{H}_I$ during the procedure. This implies that every edge of the sunflower was in $\mathcal{H}_0$ right before e was removed, i.e. that $\mathcal{H}_0$ contains a sunflower with at least s petals and kernel K. This contradicts us removing e from $\mathcal{H}_0$ at this step, so we include no such sunflower exists in any $\mathcal{H}_I$. □

Claim 28.13. *Let $\mathcal{H}$ be an r-partite r-graph and I a set such that $I \neq \text{proj}(K)$ for any K which is the kernel of a sunflower with at least s petals. Then one can decompose $\mathcal{H}$ as $\mathcal{H} = \mathcal{H}^1 \cup \cdots \mathcal{H}^{r(s-1)}$ such that $I \notin \mathcal{I}(\mathcal{H}^j)$ for all j.*

Proof. Note that for each edge $e \in \mathcal{H}$, there exists a unique set $K \subseteq V$ with $\text{proj}(K) = I$. For each set of this form, let $\mathcal{H}(K)$ be its link hypergraph, i.e. $H(K) = \{e \setminus K : K \subseteq e \in H\}$. By hypothesis, none of the $\mathcal{H}(K)$ sets contain a matching of size s (since this translates to a sunflower of size s in $\mathcal{H}$ with kernel K). It is not difficult to see that one can decompose each $(r - |K|)$-graph $\mathcal{H}(K)$ into at most $(r - |K|)(s-1) \le r(s-1)$ intersecting hypergrpahs $\mathcal{H}^1(K), \ldots, \mathcal{H}^{r(s-1)}(K)$ (e.g. by taking a largest matching M in $\mathcal{H}(K)$ and then assigning edges to $\mathcal{H}^i(K)$ if they contain the ith vertex which is contained in an edge of M). Let $\mathcal{H}^i[K] = \{e' \cup K : e' \in \mathcal{H}^i(K)\}$ and let $\mathcal{H}^i = \bigcup_K \mathcal{H}^i(K)$. Note that this decomposes $\mathcal{H}$, and that $K \neq e \cap f$ for any K with $\text{proj}(K) = I$ and $e, f \in H^i$ (as this would imply $e \setminus K, f \setminus K \in \mathcal{H}^i(K)$, and hence e, f contain an additional vertex since $\mathcal{H}^i(K)$ is intersecting). □

By repeatedly applying the above two times a bounded number of times, one can decompose $\mathcal{H}$ as $\bigcup \mathcal{H}_i$ where each $\mathcal{H}_i$ satisfies (2) and (3). Taking the largest of these hypergraphs gives the desired result.

28.2 Other Hypergraph Approximations

The results of this section are very similar in spirit to those of Section 10. To close this section, we briefly compare and contrast these results.

By using the Delta-system method (and more precisely the semi-lattice intersection lemma), we are able to conclude that many of the links of each $e \in \mathcal{H}'$ contains large sunflowers. In contrast, the spreadness of Theorem 10.2 not only gives us large sunflowers, but the stronger fact that a random partitioning of our vertex set is likely to give a large sunflower (which is a more robust condition). Moreover, the "error term" of Theorem 10.2 is typically much smaller than that of Lemma 28.4 (which only approximates a constant proportion of $\mathcal{H}$, which is in some sense necessary since it can only perfectly approximate r-partite r-graphs). On the other hand, the approximating hypergraph of Theorem 10.2 is more "complex", in the sense that it is not just a hypergraph on $[r]$ but all of $V(\mathcal{H})$, and moreover here you lose the homogeneity of Lemma 28.4 which implies that every two edges have the same intersection pattern. In conclusion, the two methods are overall incomparable to each other, with each finding different uses in different situations.

We also note that there is another famous hypergraph approximation called the Junta method, which was developed by Dinur and Friedgut [51]. This roughly says that if $\mathcal{H}$ is an intersecting hypergraph, then there exists a hypergraph $\mathcal{J}$ on a small set of vertices J of $\mathcal{H}$ such that almost every edge $e \in \mathcal{H}$ has $e \cap J \in \mathcal{J}$. We omit going into this in detail and refer the reader to [51].

Finally, we emphasize a fundamental weakness in all of these approaches, which is that they only work when n is quite large. In many cases we're okay with this, but for some results like the t-intersecting Erdős-Ko-Rado theorem, it is of interest in nailing down the exact dependency on n. In cases such as these a more careful argument is needed.

29 Counting Homomorphisms

Many problems in Extremal combinatorics center around showing the existence of a structure F in a large object, and for this it is often useful to establish a counting result for substructures related to F. For example, the standard proof of the Kővári-Sós-Turán relies on counting pairs (v, S) where $S \subseteq N(v)$ is a set of size s and then uses that the number of such pairs is large in dense graphs but not too large if the graph is $K_{s,t}$-free.

In this chapter we focus on utilizing counting results for graph homomorphisms, which typically involves two steps. For the first and generally easier step, we show that if G is a dense graph, then there are many homomorphisms from F to G, e.g. by showing that F is Sidorenko. Second, we show that the number of "degenerate" (i.e. non-injective) homomorphisms is strictly smaller than the number of homomorphisms guaranteed in the first part. This implies that there exists an injective homomorphism from F to G, and hence a copy of F in G as desired. This strategy outlined above was largely pioneered by Janzer [91] in the context of studying rainbow Turán problems, and we will be using these problems as the guiding examples throughout this chapter.

29.1 Rainbow Turán Numbers

Given a set of graphs F and an integer n, we define the *rainbow Turán number* $\text{ex}^*(n, F)$ to be the maximum number of edges that a properly colored n-vertex graph G can have without containing any rainbow copy of F (i.e. a copy of F where every edge has a distinct color in G), and we similarly define $\text{ex}^*(n, \mathcal{F})$ for a family of graphs $\mathcal{F}$. For example, we always have $\text{ex}(n, F) \le \text{ex}^*(n, F)$ since taking any proper coloring of an F-free graph G with $\text{ex}(n, F)$ edges gives a graph which is rainbow F-free. This simple inequality implies that for any given F, proving lower bounds is (potentially) easier for $\text{ex}^*(n, F)$ compared to the classic Turán problem and that upper bounds are (potentially) harder.

The rainbow Turán problem was originally introduced by Keevash, Mubayi, Sudakov, and Verstraëte [99] where much of their focus was on studying cycles. In the classic Turán setting, a result of Bondy and Simonovits [26] shows $\text{ex}(n, C_{2k}) = O(n^{1+1/k})$. This bound is only known to be tight only for $k \in \{2, 3, 5\}$ despite a large amount of work being done on trying to show this bound holds for k in general. This lower bound problem becomes significantly easier in the rainbow setting, and in particular it was shown in [99] that $\text{ex}^*(n, C_{2k}) = \Omega(n^{1+1/k})$ holds for all k.

While this lower bound for even cycles becomes significantly easier in the rainbow setting, its corresponding upper bound becomes significantly harder. Indeed, the authors of [99] were only being able to show the matching upper bound of $\text{ex}^*(n, C_{2k}) = O(n^{1+1/k})$ for the cases $k = 2, 3$. In particular, the fact that the original proof of Bondy and Simonovits can not be adapted to prove this result shows that the method of the original proof is not "robust" enough to handle the extra constraints imposed by being in the rainbow setting. This barrier was eventually overcome by Janzer [91] (see Theorem 29.4) who showed $\text{ex}^*(n, C_{2k}) = O(n^{1+1/k})$ for all k through the use of his homomorphism counting method.

Another frustrating simple case of the rainbow Turán problem is that of the set of all cycles $\mathcal{C}$. Of course, in the classical Turán setting it is easy to show $\text{ex}(n, \mathcal{C}) = n-1$, but this problem becomes surprisingly difficult in the rainbow setting. It was shown in [99] that $\mathrm{ex}^*(n, \mathcal{C}) = \Omega(n \log n)$ by taking G to be a hypercube whose edges xy are colored by the bit for which x, y differ on and they conjectured that this bound is best possible. The first big progress towards this conjecture was done by Das, Lee, and Sudakov [46] who showed an upper bound of roughly $ne^{\sqrt{\log n}}$. This was then substantially improved by Janzer [91] who showed an upper bound of $O(n(\log n)^4)$ using his homomorphism counting argument. Other homomorphism based approaches were later used independently by Janzer and Sudakov [?] and by Kim, Lee, Liu, and Tran [100] to give an upper bound of $O(n(\log n)^2)$. Finally, this problem was essentially solved (through non-homomorphic means) by Alon, Bucić, Sauermann, Zakharov, and Zamir [3] who showed an upper bound of $n(\log n)^{1+o(1)}$.

In what follows we look at two different homomorphism based proofs of the upper bound $\mathrm{ex}^*(n, C_{2k})$. We will begin by going through Janzer's original short proof of the result, after which we will sketch out a later proof found by [100] whose approach is longer but perhaps more intuitive.

29.2 The First Proof

As noted at the start of the chapter, the key step in Janzer's homomorphism counting method is to upper bound the number of "degenerate" homomorphisms. In the classic Turán setting "degenerate" typically just means non-injective, but in e.g. the rainbow Turán setting we will need to additionally say that any cycle which has repeated colors is also degenerate. More broadly, we will prove the following lemma which works for a fairly abstract notion of "degeneracy" defined in terms of a binary relation $\sim$. Here and throughout we abuse notation slightly and write $\hom(C_2, G) := \hom(K_2, G)$ and $\hom(C_0, G) := \hom(K_1, G)$. For convenience we will often denote homomorphisms of C_{2k} to G by sequences $(x_1, \ldots, x_{2k})$ such that $x_i x_{i+1} \in E(G)$ for all $1 \le i \le 2k$ with indices written cyclically.

Lemma 29.1. *Let G be a graph with maximum degree Δ and let $\sim$ be a symmetric binary relation on the vertices of G such that for every $(u, v) \in V(G)^2$ and $w \in V(G)$, the vertex w has at most s neighbors z satisfying $(u, v) \sim (z, w)$.*

Let $\hom_{\deg}(C_{2k}, G)$ denote the number of homomorphisms $(x_1, \ldots, x_{2k})$ from C_{2k} to G with $(x_i, x_{i+1}) \sim (x_j, x_{j+1})$ for some $i \ne j$. Then for all $k \ge 2$ we have

$$\hom_{\deg}(C_{2k}, G) \le 16k(ks\Delta \hom(C_{2k-2}, G) \hom(C_{2k}, G))^{1/2}.$$

For example, in the setting of finding rainbow C_{2k}'s in a properly edged colored graph G, we will define $(x, y) \sim (z, w)$ whenever xy, zw are edges which either have the same color or $x = z$, as in this case a rainbow cycle is exactly a non-degenerate homomorphism from C_{2k}.

Proof. The main idea for this proof is to partition the set of degenerate cycle homomorphisms $(x_1, \ldots, x_{2k})$ based on how many walks there are from, say, x_1 to x_{k+2}, and then use different types of arguments depending on how large these quantities are. To this end, for each $\ell \ge 1$ and $u, v \in V(G)$, we define $w_\ell(u, v)$ to be the number of walks of length ℓ from u to v (equivalently, this is the number of homomorphisms of a path of length ℓ which has u, v as its endpoints). For all $r, t \geq 1$, let $\gamma_{r,t}$ denote the number of degenerate cycle homomorphisms $(x_1, \ldots, x_{2k})$ which have:

(a) $2^{r-1} \leq w_{k-1}(x_1, x_{k+2}) < 2^r$,

(b) $2^{t-1} \leq w_k(x_2, x_{k+1}) < 2^t$, and

(c) $(x_1, x_2) \sim (x_i, x_{i+1})$ for some $2 \leq i \leq k+1$.

Observe that every degenerate cycle homomorphisms $(y_1, \ldots, y_{2k})$ can be cyclically shifted to some $(x_1, \ldots, x_{2k})$ with $(x_1, x_2) \sim (x_i, x_{i+1})$ for some $2 \leq i \leq k+1$ which is necessarily counted by some $\gamma_{r,t}$, from which we see that

$$\hom_{\deg}(C_{2k}, G) \leq 2k \sum_{r,t} \gamma_{r,t}, \tag{24}$$

so it suffices to bound this sum.

We introduce two related parameters to help bound the $\gamma_{r,t}$ terms: let α_r denote the number of walks $(y_1, \ldots, y_k)$ such that $2^{r-1} \leq w_{k-1}(y_1, y_k) < 2^r$, and similarly let β_t denote the number of walks $(z_1, \ldots, z_{k+1})$ such that $2^{t-1} \leq (z_1, z_{k+1}) < 2^t$. Since each $C_{2\ell}$ homomorphism consists of choosing a walk of length ℓ followed by another walk of length ℓ between its two endpoints, we see that

$$\hom(C_{2k-2}, G) \geq \sum_r \alpha_r 2^{r-1},$$

$$\hom(C_{2k}, G) \geq \sum_t \beta_t 2^{t-1}.$$

Claim 29.2. *For all r, t, we have*

$$\gamma_{r,t} \leq \alpha_r \cdot \Delta \cdot 2^t,$$

$$\gamma_{r,t} \leq \beta_t \cdot ks \cdot 2^r.$$

Proof. Each cycle homomorphism $(x_1, \ldots, x_{2k})$ counted by $\gamma_{r,t}$ can be identified as follows: choose the walk $(x_1, x_{2k}, x_{2k-1}, \ldots, x_{k+2})$, then the neighbor x_2 of x_1, then the walk $(x_2, \ldots, x_{k+2})$. Note that the number of choices for the first walk is α_r by definition of α_r and $\gamma_{r,t}$, the number of choices for x_2 is trivially at most Δ, and the number of choices for the last walk is at most 2^t again by definition of $\gamma_{r,t}$. This gives the first bound.

One can alternatively identify the homomorphisms as follows: choose the walk $(x_2, \ldots, x_{k+2})$, then x_1, then the walk $(x_1, x_{2k}, x_{2k-1}, \ldots, x_{k+2})$. As before there are β_t choices for the first walk. Crucially, because we must choose x_1 so that $(x_1, x_2) \sim (x_i, x_{i+1})$ for some $2 \leq i \leq k+1$ (with these (x_i, x_{i+1}) edges fixed because of the order in which we specified our walks), the total number of choices for x_1 is at most ks by hypothesis of $\sim$. Finally, the last walk is chosen in at most 2^r ways by definition of $\gamma_{r,t}$, giving the result. □

We will choose which bound in Claim 29.2 to use for a given $\gamma_{r,t}$ depending on the relative size of r and t. To this end, we fix some cutoff value q to be determined later and observe by

Claim 29.2 that

$$\begin{aligned}\sum_{r,t}\gamma_{r,t} &= \sum_{r,t:t<r+q}\gamma_{r,t} + \sum_{r,t:t\geq r+q}\gamma_{r,t}\\ &\leq \Delta\sum_{r,t:t<r+q}2^t\alpha_r + ks\sum_{r,t:t\geq r+q}2^r\beta_t\\ &\leq 2^q\Delta\sum_{r\geq 1}2^r\alpha_r + ks2^{-q+1}\sum_{r,t:t\geq r+q}2^t\beta_t\\ &\leq 2^{q+1}\Delta\hom(C_{2k-2},G) + ks2^{-q+2}\hom(C_{2k},G),\end{aligned} \tag{25}$$

where the third line used that, for example, we have for any fixed r that $\sum_{t:t<r+q}2^t \leq 2^{r+q}$; and the last line used the inequalities just before Claim 29.2. We can optimize (25) by choosing q so that these two terms are close to equal; say by taking q to be the unique integer with

$$\left(\frac{ks\cdot\hom(C_{2k},G)}{\Delta\hom(C_{2k-2},G)}\right)^{1/2} \leq 2^q < 2\left(\frac{ks\cdot\hom(C_{2k},G)}{\Delta\hom(C_{2k-2},G)}\right)^{1/2}.$$

In this case each of the two terms (25) can be bounded by $4(ks\Delta\hom(C_{2k-2},G)\hom(C_{2k},G))^{1/2}$, which combined with (24) gives

$$\hom_{\deg}(C_{2k},G) \leq 2k\cdot 8(ks\Delta\hom(C_{2k-2},G)\hom(C_{2k},G))^{1/2},$$

completing the proof. □

As an aside, it might seem more natural in the proof to define e.g. α_r to count the number of walks with $w_{k-1}(y_1,y_k) = r$ rather than define it in terms of a dyadic partition. If one makes this change, then most of the present proof will go through (with slightly different notation) until one tries to bound the term

$$\sum_{r,t:t<qr}t\alpha_r.$$

In this case, upon fixing any value of r the sum over the t term becomes something like r^2, which ends up killing the proof.

We want to use Lemma 29.1 to show that if G has many edges, then it contains a non-degenerate C_{2k}. For this, we need to show that (a) $\hom(C_{2k-2},G)$ is not too large in terms of $\hom(C_{2k},G)$ and (b) $\hom(C_{2k},G)$ is large whenever G has many edges (and in particular, larger than our upper bound for the number of degenerate homomorphisms). This will be accomplished through the following.

Proposition 29.3. *Let G be an n-vertex graph and $k \geq 2$ an integer.*

(a) We have

$$\hom(C_{2k-2},G) \leq n^{1/k}\hom(C_{2k},G)^{1-1/k}.$$

(b) We have

$$\hom(C_{2k},G) \geq \left(\frac{2e(G)}{n}\right)^{2k}.$$

Note that (b) is equivalent to saying that C_{2k} satisfies Sidorenko's conjecture, which is a typical prerequisite for being able to apply this method in practice. We will postpone the proof of Proposition 29.3 for the moment and show how it implies our main result for this subsection.

Theorem 29.4. *For all $k \geq 2$, we have $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$.*

Proof. For ease of presentation we only prove the following weaker version of the theorem: if G is an n-vertex graph whose edges are properly colored such that G does not contain a rainbow C_{2k} and is such that G is *regular*, then $e(G) = O(n^{1+1/k})$. We note that one can drop the extra regularity condition by using standard results that let one pass to "almost regular" subgraphs of G (see for example [94, Proposition 2.7]), but we restrict ourselves to this slightly restricted setting in order to make our arguments involving the main ideas as clear as possible.

Let Δ denote the (maximum) degree of G and define a symmetric binary relation $\sim$ on $V(G)^2$ by having $(x, y) \sim (z, w)$ if xy, zw are both edges in G and either (a) xy, zw have the same color in G or (b) $x = z$. Observe that for any given (u, v) and w, there are at most 2 neighbors z of w such that $(u, v) \sim (z, w)$ because G is properly colored. As such, Lemma 29.1 applies and shows that

$$\hom_{\deg}(C_{2k}, G) \leq 16k(2k\Delta \hom(C_{2k-2}, G) \hom(C_{2k}, G))^{1/2} \leq 32k^{3/2}\Delta^{1/2}n^{1/2k} \hom(C_{2k}, G)^{1-1/2k},$$

where this last inequality used Proposition 29.3(b). On the other hand, Proposition 29.3(a) together with G being Δ-regular implies $\hom(C_{2k}, G) \geq \Delta^{2k}$, i.e. that $\hom(C_{2k}, G)^{1/2k} \geq \Delta$. Using this in the expression above gives

$$\hom_{\deg}(C_{2k}, G) \leq 32k^{3/2}\Delta^{-1/2}n^{1/2k} \hom(C_{2k}, G).$$

Note that if $\Delta > 2^{12}k^3 n^{1/k}$ then this implies $\hom(C_{2k}, G) > \hom_{\deg}(C_{2k}, G)$, i.e. that G contains a C_{2k} homomorphism $(x_1, \ldots, x_{2k})$ such that $(x_i, x_{i+1}) \not\sim (x_j, x_{j+1})$ for any $i \neq j$. By definition of $\sim$ this would imply G contains a rainbow cycle, a contradiction. We thus must have $\Delta = O(n^{1/k})$, i.e. $e(G) = O(n^{1+1/k})$ as desired. $\square$

The contradiction coming from $\Delta \gg k^3 n^{1/k}$ at the end of this proof implies that if G is a properly-colored *regular* graph with at least $\Omega(n(\log n)^3)$ edges, then G contains a rainbow cycle, namely one of length $\Theta(\log n)$, and Janzer was able to get around this regularity condition at the cost of a $\log n$ to prove his upper bound $\mathrm{ex}^*(n, \mathcal{C}) = O(n(\log n)^4)$ mentioned previously.

It remains to prove Proposition 29.3, which we will do through the following basic spectral graph theory result.

Lemma 29.5. *If G is an n-vertex graph and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of its adjacency matrix A, then $\hom(C_{2\ell}, G) = \sum \lambda_i^{2\ell}$ for all $\ell \geq 0$.*

Proof. It is easy to check that entrywise we have $A^t_{u,v} = w_t(u, v)$ for all $u, v \in G$ and $t \in \mathbb{N}$. In particular, the total number of closed walks of length 2ℓ equals $\sum_u A^{2\ell}_{u,u} = \mathrm{Tr}(A^{2\ell})$. The result follows since the eigenvalues for $A^{2\ell}$ are $\lambda_1^{2\ell}, \ldots, \lambda_n^{2\ell}$. $\square$

We now move onto prove our technical homomorphism inequalities.

Proof of Proposition 29.3. We begin by proving (b). By the lemma above, we in particular have $\hom(C_{2k}, G) \geq \lambda_1^{2k}$ where λ_1 is the largest eigenvalue of the adjacency matrix A, so it will suffice to show $\lambda_1 \geq \frac{2e(G)}{n}$. To this end, we observe by the Raleigh quotient that

$$\lambda_1 = \max_v \frac{v^T A v}{v^T v},$$

and taking v to be the all 1 vector exactly gives $\lambda_1 \geq \frac{2e(G)}{n}$, proving the result.

For (a), we first claim that for any $\ell \geq 2$ we have

$$\frac{\hom(C_{2\ell-2}, G)}{\hom(C_{2\ell-4}, G)} \leq \frac{\hom(C_{2\ell}, G)}{\hom(C_{2\ell-2}, G)}. \tag{26}$$

Indeed, this is equivalent to claiming $\hom(C_{2\ell}, G)\hom(C_{2\ell-4}, G) \geq \hom(C_{2\ell-2}, G)^2$, and by the lemma above this is equivalent to showing

$$\left(\sum_i \lambda_i^{2\ell}\right)\left(\sum_i \lambda_i^{2\ell-4}\right) \geq \left(\sum_i \lambda_i^{2\ell-2}\right)^2,$$

which follows from the Cauchy-Schwartz inequality (applied with $x_i = \lambda_i^{\ell}$ and $y_i = \lambda_i^{\ell-2}$), proving the claim.

By iteratively applying (26), we find that

$$\frac{\hom(C_{2k}, G)}{\hom(C_0, G)} = \frac{\hom(C_{2k}, G)}{\hom(C_{2k-2}, G)} \cdots \frac{\hom(C_2, G)}{\hom(C_0, G)} \leq \left(\frac{\hom(C_{2k}, G)}{\hom(C_{2k-2}, G)}\right)^k.$$

Rearranging this inequality and using $\hom(C_0, G) = n$ gives the desired result. □

29.3 An Alternative Approach

Here we discuss an alternative way to prove $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$ due to [100] whose general approach is perhaps more intuitive than Janzer's original proof. Roughly speaking, their main result shows that if a large proportion of the homorphisms from C_{2k} are degenerate, then in fact a large proportion of them are "extremely" degenerate, i.e. come from mapping all of the vertices of even index in C_{2k} to the same vertex in G, which is the same as a homomorphism from $K_{1,k}$. At a high level, they prove this result by considering a sequence of graphs $F_0, \ldots, F_{k-1}$ of increasing "degeneracy" starting with $F_0 = C_{2k}$ and ending with $F_{k-1} = K_{1,k}$, and then from here iteratively show that $\hom(F_{s+1}, G) \approx \hom(F_s, G)$ for all s, concluding with $\hom(K_{1,k}, G) \approx \hom(C_{2k}, G) \approx \hom_{\mathrm{deg}}(C_{2k}, G)$ since we assumed a large proportion of homomorphisms are degenerate.

More precisely, the key homomorphism inequality proved in [100] is the following[1], where here we again denote homomorphisms of C_{2k} to G by sequences $(x_1, \ldots, x_{2k})$ such that $x_i x_{i+1} \in E(G)$ for all $1 \leq i \leq 2k$.

[1]Strictly speaking they prove this bound only under the hypothesis that G has no rainbow C_{2k}, i.e. when $D = 1$, but it is easy to derive our formulation from their proof.

Lemma 29.6. *Given a graph G with a proper edge coloring, let $\hom_{\deg}(C_{2k}, G)$ denote the number of C_{2k} homomorphisms $(x_1, \ldots, x_{2k})$ where some vertex or edge color is repeated. If there exists $D > 0$ such that $\hom(C_{2k}, G) \leq D \cdot \hom_{\deg}(C_{2k}, G)$, then*

$$\hom_{\deg}(C_{2k}, G) \leq (4k^2 D)^{k-1} \hom(K_{1,k}, G).$$

Note that we trivially have $\hom(K_{1,k}, G) \leq \hom_{\deg}(C_{2k}, G)$ since each $K_{1,k}$ homomorphism can be viewed as a homomorphism of C_{2k} where each vertex of even index gets mapped to the same vertex in G, so this lemma says that this trivial bound is close to tight whenever a large proportion of the C_{2k} homomorphisms are degenerate. This bound is particularly useful since $\hom(K_{1,k}, G)$ depends only on the degree sequence of G and hence is easy to bound.

Sketch of Proof. While the high level ideas of this proof are rather clean, the exact computations are somewhat lengthy. As such, we will omit most of the computations here and focus mostly on the core ideas running through the argument.

As noted above, we can essentially think of $K_{1,k}$ as counting the homomorphisms of C_{2k} where all the vertices of even index map to the same vertex in G, and ultimately we want to show that such "very degenerate" homomorphisms are close to counting the total number of degenerate homomorphisms. We will do this by "improving" the degeneracy of these "very degenerate" homomorphisms one step at a time while making sure that the total count does not change much as we do this.

More precisely, for $0 \leq s \leq k-1$ we let $\mathcal{O}_s$ denote the set of cycle homomorphisms $(x_1, \ldots, x_{2k})$ such that $x_1 = x_3 = \cdots = x_{2s+1}$. Essentially then, $\mathcal{O}_s$ corresponds to the homomorphisms of the graph which is defined by taking a $K_{1,s}$ and attaching a C_{2k-2s} to its center vertex. In particular, $|\mathcal{O}_0| = \hom(C_{2k}, G)$ and $|\mathcal{O}_{k-1}| = \hom(K_{1,k}, G)$.

Note that we trivially have $|\mathcal{O}_s| \leq |\mathcal{O}_{s-1}|$ for all s, and we would like to show that this is close to an equality. As a first step, we show that the successive ratios $\frac{|\mathcal{O}_{s-1}|}{|\mathcal{O}_s|}$ are decreasing in s, from which we will only need to bound the first ratio to get an effective bound. More precisely, we show the following.

Claim 29.7. *For all $1 \leq s \leq k-2$, we have $|\mathcal{O}_s|^2 \leq |\mathcal{O}_{s-1}||\mathcal{O}_{s+1}|$.*

Proof. We prove only the case when s is odd, the s even case being similar but requiring a bit more care. Define a star walk of length 2ℓ to be a walk of the form $(u_1, \ldots, u_{2\ell+1})$ such that $u_1 = u_3 = \cdots = u_{2\ell+1}$, and let $\sigma_{2\ell}(x)$ denote the number of such walks with $u_1 = x$.

Observe that each homomorphism in $\mathcal{O}_s$ can be counted as follows: (1) fix $x := x_1 = x_3 = \cdots = x_{2s+1}$ and $z = x_{k+s+2}$, (2) choose a walk $(x_{k+s+2}, x_{k+s+3}, \ldots, x_{2k}, x_1)$ of length $k-s-1$ from z to x, (3) choose a star walk $(x_1, x_2, \ldots, x_{s+2})$ of length $s+2$ from x back to itself, (4) choose a walk $(x_{k+s+2}, x_{k+s+1}, \ldots, x_{2s+1})$ from z to x, (5) choose a star-walk $(x_{2s+1}, x_{2s}, \ldots, x_{s+2})$ from x to itself. In total then, this implies

$$|\mathcal{O}_s| = \sum_{x,z} w_{k-s-1}(z,x)\sigma_{s+1}(x) w_{k-s+1}(z,x)\sigma_{s-1}(x),$$

so by the Cauchy-Schwartz inequality we have

$$|\mathcal{O}_s|^2 \le \left(\sum_{x,z} w_{k-s-1}(z,x)^2 \sigma_{s+1}(s)^2\right)\left(\sum_{x,z} w_{k-s+1}(z,x)^2\sigma_{s-1}(s)^2\right) = |\mathcal{O}_{s+1}||\mathcal{O}_{s-1}|,$$

where the last inequality follows a similar logic as above. □

Note that this claim implies

$$\frac{\hom(C_{2k},G)}{\hom(K_{1,k},G)} = \frac{|\mathcal{O}_0|}{|\mathcal{O}_{k-1}|} = \prod_{s=0}^{k-2}\frac{|\mathcal{O}_s|}{|\mathcal{O}_{s+1}|} \le \left(\frac{\hom(C_{2k},G)}{|\mathcal{O}_1|}\right)^{k-1}, \tag{27}$$

and since $\hom(C_{2k},G)$ and $\hom_{\deg}(C_{2k},G)$ differ by a constant factor, it will suffice to show $\hom_{\deg}(C_{2k},G)$ differs by a constant factor from $|\mathcal{O}_1|$, i.e. those C_{2k} homomorphisms with $x_1 = x_3$.

Note that $|\mathcal{O}_1|$ is a lower bound for $\hom_{\deg}(C_{2k},G)$, but there exist many more homomorphisms, e.g. those with $x_1 = x_i$ for some $i > 3$, and we might think of such homomorphisms as being "more degenerate" the larger i is. As we did for the $\mathcal{O}_s$ sets, we will be able to make small moves to show that these "more degenerate" homomorphisms behave similarly to $\mathcal{O}_1$. To this end, we define $\mathcal{U}_s$ to be the set of homomorphisms $(x_1,\ldots,x_{2k})$ with $x_1 = x_{s+2}$, noting that $\mathcal{U}_1 = \mathcal{O}_1$.

Claim 29.8. *For all $1 \le s \le k-1$, we have $|\mathcal{U}_s|^2 \le |\mathcal{U}_1| \cdot |\mathcal{U}_{2s-1}|$. In particular, we have $|\mathcal{U}_s| \le |\mathcal{U}_1|$ for all $1 \le s \le 2k-3$.*

The first half of the claim is proved similarly to the previous claim by decomposing the homomorphism into certain walks and then using Cauchy-Schwartz; we omit the details. The second half follows after observing that $|\mathcal{U}_s| = |\mathcal{U}_{2k-2-s}|$ by symmetry to utilize e.g. $|\mathcal{U}_{k-1}|^2 \le |\mathcal{U}_1| \cdot |\mathcal{U}_{2k-3}| = |\mathcal{U}_1|^2$.

Observe that if a homomorphism $(x_1,\ldots,x_{2k})$ has $x_i = x_j$ for some $i \ne j$, then we can cyclically rotate this so that $i = 1$ and $3 \le j \le k+1$. As such, the claim above implies that the number of degenerate homomrphisms with $x_i = x_j$ for some $i \ne j$ is at most

$$2k\sum_{s=1}^{k-1}|\mathcal{U}_s| \le 2k^2|\mathcal{U}_1| = 2k^2|\mathcal{O}_1|.$$

The degenerate homomorphisms with x_ix_{i+1} colored the same as x_jx_{j+1} can be dealt with in a similar way: here we define $\mathcal{F}_s$ to be those homomorphisms with x_1x_2 colored the same as $x_{s+1}x_{s+2}$ (noting $\mathcal{F}_1 = \mathcal{U}_1 = \mathcal{O}_s$). The exact same claim and bounds for $\mathcal{U}_s$ continue to hold for $\mathcal{F}_s$ with similar proofs, in total giving the bound

$$\hom_{\deg}(C_{2k},G) \le 4k^2|\mathcal{O}_1|,$$

or equivalently

$$\hom(C_{2k},G) \le 4k^2D|\mathcal{O}_1|$$

by the hypothesis of the lemma. This combined with (27) gives the desired result. □

This lemma quickly gives another proof of Theorem 29.4 showing that $\mathrm{ex}^*(n, C_{2k}) = O(n^{1+1/k})$.

Alternative Proof of Theorem 29.4. As in our first proof, we will for simplicity work with a properly colored n-vertex graph G which is *regular*, say of degree Δ, and which contains no rainbow C_{2k}. In the language of Lemma 29.6, not containing a rainbow C_{2k} means $\hom_{\deg}(C_{2k}, G) = \hom(C_{2k}, G)$. This implies $D = 1$ in Lemma 29.6, so we conclude that

$$\hom(C_{2k}, G) = \hom_{\deg}(C_{2k}, G) \leq (4k^2)^{k-1} \hom(K_{1,k}, G) = (4k^2)^{k-1}\Delta^k n.$$

On the other hand, we have from Proposition 29.3(b) that $\hom(C_{2k}, G) \geq \Delta^{2k}$, so in total we find $\Delta \leq 4k^2 n^{1/k}$ as desired. $\square$

Similar to our comment after the original proof of Theorem 29.4, this proof shows that if G is a properly colored regular graph without any rainbow cycle then $e(G) = O(n(\log n)^2)$, and by utilizing a more careful regularization argument than that of Janzer they were able to show $\mathrm{ex}^*(n, \mathcal{C}) = O(n(\log n)^2)$. This same upper bound $\mathrm{ex}^*(n, \mathcal{C}) = O(n(\log n)^2)$ was also obtained independently by Janzer and Sudakov [92] who developed a spiritually similar argument for the purpose of bounding Turán numbers of certain bipartite graphs F such as hypercubes. As in the present writeup, their guiding principle was to show that if a large proportion of a graphs F homomorphisms are degenerate, then most degenerate homomorphisms are in fact "very degenerate" in the sense that they can bounded by the number of star homomorphisms (i.e. by those F homomorphisms which map one of the parts of F to a single vertex).

30 Expansion and α-maximality

This needs to be heavily rewritten to incorporate improvements to the bounds and the claimed simplification mentioned at the end of "Towards the Erdős-Gallai Cycle Decomposition conjecture"

One of the nice features of random graphs is that they have good expansion properties; e.g. any set of vertices $B \subseteq V(G_{n,p})$ is likely to have about $pn|B|$ edges leaving B provided B is not too small. It is too much to ask that a graph has such strong expansion properties in general, but it is often the case that one can find subgraphs of arbitrary graphs which have reasonable expansion properties. There are many techniques in the field that achieve this end. The focus on this chapter will be an approached introduced by Tomon [142] which has the advantage of being both simple to state and powerful in applications.

Definition 5. Given a real number α, we say that a graph G is α-maximal if $e(G)/v(G)^{1+\alpha} = \max_{H\subseteq G} e(H)/v(H)^{1+\alpha}$. Equivalently, this says that if $e(G) = \gamma \cdot v(G)^{1+\alpha}$, then $e(H) \le \gamma \cdot v(H)^{1+\alpha}$ for all $H \subseteq G$.

Observe that every graph has an α-maximal subgraph.

The motivation for this definition is that often in extremal graph theory, one wants to prove that graphs with $e(G) \ge \gamma v(G)^{1+\alpha}$ contain some desired structure. If such a result were true, then in particular any α-maximal subgraph of G must contain this structure, so being α-maximal is essentially the hardest case that one can consider. Moreover, it turns out that by reducing to α-maximal graphs, one gains a lot of nice expansion properties. Here and throughout this section we let $N(B)$ be the set of vertices $y \notin N(B)$ which are adjacent to a vertex in B, and we let $d(G)$ denote the average degree of G.

Proposition 30.1. *Let G be an n-vertex α-maximal graph with $\alpha \in (0,1]$ and $d(G) = \gamma n^\alpha$, and let $B \subseteq V(G)$ be such that $|B| \le n/2$.*

(i) If G is non-empty, then $\gamma \ge \frac{1}{2}$.

(ii) The minimum degree of G is at least $\frac{1}{2}d(G) = \frac{1}{2}\gamma n^\alpha$.

(iii) We have $e(B, N(B)) \ge \frac{1}{4}\gamma n^\alpha |B|(1 + \alpha - (2|B|/n)^\alpha)$.

(iv) We have $|N(B)| > |B|((1 + \frac{1}{2}\alpha)(\frac{n}{2|B|})^{\alpha/(1+\alpha)} - 1)$.

The main benefit of (i) is that the bound is an absolute constant independent of α. Condition (ii) is obviously convenient to have. Note that in $G_{n,p}$ with $p = \gamma n^\alpha/n$, we have $\mathbb{E}[[e(B, N(B))] \approx \gamma n^\alpha |B|$ as long as $|B| \le n/2$, so the level of expansion in (iii) is about as much as we could hope for. The bound for (iv) is roughly $(n^\alpha|B|)^{1/(1+\alpha)}$, which is best possible when $|B| \approx n$ **Though beyond this I don't have much intuition for why this is a reasonable condition to shoot for.**

Proof. For (i), taking $H \subseteq G$ to be a single edge implies $d(H)/v(H)^\alpha = 2^{-\alpha} \ge \frac{1}{2}$, so the same bound holds for G.

For (ii), let v be a vertex of minimum degree δ and let $H = G - v$. We have $d(H)/v(H)^\alpha \leq d(G)/v(G)^\alpha$ by definition of α-maximality, which is equivalent to

$$\frac{d(G)n - 2\delta}{(n-1)^{1+\alpha}} \leq \frac{d(G)}{n^\alpha}.$$

This implies

$$\frac{1}{2}d(G)(n - \frac{(n-1)^{1+\alpha}}{n^\alpha}) \leq \delta,$$

showing $\delta \geq \frac{1}{2}d(G)$.

For (iii) and (iv), let $C = V(G) \setminus B$. With this we have

$$\begin{aligned} e(B, N(B)) &= e(G[B \cup C]) - e(G[B]) - e(G[C]) = \frac{1}{2}\gamma(|B| + |C|)^{1+\alpha} - e(G[B]) - e(G[C]) \\ &\geq \frac{1}{2}\gamma|C|^{1+\alpha}((1 + |B|/|C|)^{1+\alpha} - \frac{1}{2}\gamma|B|^{1+\alpha} - \frac{1}{2}\gamma|C|^{1+\alpha}. \end{aligned}$$

Using $(1 + |B|/|C|)^{1+\alpha} \geq 1 + (1 + \alpha)|B|/|C|$ and that $|C| \geq (n/2)^\alpha \geq \frac{1}{2}n^\alpha$, we find that this is at least

$$\frac{1}{2}\gamma(1 + \alpha)|B||C|^\alpha - \frac{1}{2}\gamma|B|^{1+\alpha} \geq \frac{1}{2}\gamma|B|(\frac{1}{2}(1 + \alpha)n^\alpha - |B|^\alpha),$$

giving (iii).

Similarly for (iv) we observe

$$e(G[B \cup N(B)]) \geq e(G[B \cup C]) - e(G[C]) \geq \frac{1}{2}\gamma(|B| + |C|)^{1+\alpha} - \frac{1}{2}\gamma|C|^{1+\alpha} \geq \frac{1}{2}\gamma(1 + \alpha)|B||C|^\alpha.$$

However, by α-maximality we have $e(G[B \cup N(B)]) \leq \frac{1}{2}\gamma(|B| + |N(B)|)^{1+\alpha}$. Combining these inequalities gives

$$|N(B)| \geq ((1 + \alpha)|B||C|^\alpha)^{1/(1+\alpha)} - |B|,$$

giving the result. □

Our main application of α-maximal graphs will be to something called rainbow Turán numbers, which were first introduced by Keevash, Mubayi, Sudakov, and Verstraëte [99]. We say that a colored graph F is *rainbow* if all of the colors of its edges are distinct. Given a set of graphs $\mathcal{F}$, we define $\mathrm{ex}^*(n, \mathcal{F})$ to be the maximum number of edges a *properly colored* n-vertex graph G can have without containing a rainbow copy of any $F \in \mathcal{F}$.

Note that $\mathrm{ex}(n, \mathcal{F}) \leq \mathrm{ex}^*(n, \mathcal{F})$ for all $\mathcal{F}$ (since we can take any extremal $\mathcal{F}$-free graph and give each edge a distinct color), and in general these two quantities can be somewhat far from each other. Indeed, let $\mathcal{C}$ denote the set of all cycles, which means $\mathrm{ex}(n, \mathcal{C}) = n - 1$. On the other hand, we have $\mathrm{ex}^*(n, \mathcal{C}) \geq n \log_2 n$ when n is a power of 2. This is because one can take G to be an n-vertex hypercube where an edge uv is colored i if u, v differ in the ith bit. It is not difficult to see that this is a proper coloring which contains no rainbow cycles.

Even though the problem of determining $\mathrm{ex}(n, \mathcal{C})$ is easy, determining $\mathrm{ex}^*(n, \mathcal{C})$ is an open and seemingly difficult problem. The first non-trivial upper bounds on $\mathrm{ex}^*(n, \mathcal{C})$ were established by Das, Lee, and Sudakov [46], and later O. Janzer [91] managed to prove $\mathrm{ex}^*(n, \mathcal{C}) = O((\log n)^4 n)$. Currently the best known upper bound is the following result due to Tomon [142].

Theorem 30.2 ([142]). *We have* $\mathrm{ex}^*(n, \mathcal{C}) = (\log n)^{2+o(1)} n$.

The main lemma we need to prove this is the following. Here a *Q-rainbow path* refers to a rainbow path which only uses colors in the set Q.

Lemma 30.3. *Given $p_c \in (0,1)$, there exists a constant C such that the following holds. Let $\lambda > C(\log\log n)^{10}$, and let G be an n-vertex α-maximal graph with proper coloring $c : E(G) \to R$ and $d(G) > C\lambda^2\alpha^{-2}n^{\alpha}$. If $Q \subseteq R$ is chosen by including each color independently and with probability p_c, then for every $v \in V(G)$, with probability at least $1 - O(\alpha^{-1}e^{-\Omega(\lambda^{1/2})})$ at least $n/3$ vertices of G can be reached by a Q-rainbow path.*

Before proving this lemma, let us first show how this implies the main result.

Proof of Theorem 30.2. Let G be an n-vertex graph with $e(G) \geq 2(\log n)^{2+\epsilon} n$ and $c : E(G) \to R$ a proper coloring, and let $\alpha = 1/\log_2(G)$ and $\lambda = (\log n)^{\epsilon/10}$. Let H be a subgraph of G maximizing $d(H)/v(H)^{\alpha}$ and $m = v(H)$. Note that H is α-maximal and $d(H) \geq d(G) \cdot (v(H)/v(G))^{\alpha} \geq \frac{1}{2}d(G)$ due to our choice of α, and this quantity is at least $C\lambda^2\alpha^{-2}m^{\alpha}$ for n sufficiently large.

Pick some $v \in V(H)$. Partition R into four parts Q_1, Q_2, Q_3, Q_4 be independently and uniformly at random assigning each color to one of these sets, and let B_i be the set of vertices that can be reached by v with a Q_i-path. By Lemma 30.3 with $p_c = 1/4$, we see that with probability at least $4/5$ we have $|B_i| \geq n/3$, so there exists some partition $Q_1, \ldots, Q_4$ such that $|B_i| \geq n/3$ holds for all i.

Note that $B_i \cap B_j \neq \emptyset$ for some $i \neq j$, and let $w \in B_i \cap B_j$. By definition this means there exist rainbow paths P_i, P_j from v to w using colors in Q_i, Q_j. Thus the union of these two paths is a rainbow graph which contains a cycle, proving the result. $\square$

It remains to prove Lemma 30.3. Given a graph G and a proper coloring $c : E(G) \to R$, define $N_{Q,\phi}(v)$ with $\phi : V(G) \to 2^{V(G)\cup R}$ to be the set of vertices w with $vw \in E(G)$, $c(vw) \in Q \setminus \phi(v)$, and $w \notin \phi(v)$. That is, the is the neighborhood if we restrict to colors in Q and forbid some set of neighbors/colors for v to use. We define $N_{Q,\phi}(B) = \bigcup_{v\in B} N_{Q,\phi}(v) \setminus B$. To prove Lemma 30.3, we show that α-maximal graphs have vertex expansion about as strong as in Proposition 30.1 even when forbidding some colors/vertices.

Lemma 30.4. *Let $p_c, \alpha \in (0,1]$, let n be a positive integer and $\lambda > 10^{10}$. Let G be an n-vertex graph, $c : E(G) \to R$ a proper edge coloring, and $B \subseteq V(G)$ such that the following hold:*

- *G is α-maximal*
- *$d := d(G) \geq \lambda(p_c \cdot \alpha)^{-1}$,*
- *$\phi : V(G) \to 2^{V(G)\cup R}$ is such that $|\phi(v)| \leq d\alpha/32$ for all $v \in V(G)$, and*
- *$2\lambda^2 p_c^{-1} < |B| < n/2$.*

Let $Q \subseteq R$ be obtained by including each color independently and with probability p_c. Then with probability at least $1 - e^{-\Omega(\lambda^{1/2})}$ we have

$$|N_{Q,\phi}(B)| \geq \frac{1}{4}|B| \min\left\{\frac{d \cdot p_c \cdot \alpha}{64\lambda^{1/2}}, \left(\frac{n}{2|B|}\right)^{\alpha/(1+\alpha)} - 1\right\}$$

Proof. Let $d = \gamma n^\alpha$, and let H be the bipartite (uncolored) graph on $B \cup N(B)$ such that $x \in B$ and $y \in N(B)$ are adjacent if $xy \in E(G)$, $y \notin \phi(x)$ and $c(xy) \notin \phi(x)$. Let $H_Q \subseteq H$ be the (random) subgraph which only includes edges xy with $c(xy) \in Q$. Thus our problem is equivalent to showing $|N_{H_Q}(B)|$ is large with high probability.

Since we're aiming something comparable to that of Proposition 30.1 (namely, this is basically what we get when the right term in the lemma achieves the minimum), one might try to just naively replicate that proof. This almost works, but to get things to occur with high probability we need the vertices of $N_H(B)$ to have large degrees. To this end, let $S \subseteq N_H(B)$ be the vertices w such that $|N_H(w) \cap B| \geq \lambda^{1/2} p_c^{-1} =: \Delta$, and let $T = N_G(B) \setminus S$.

Claim 30.5. *If $e_G(B, T) \leq d\alpha|B|/16$, then the result follows.*

Note that the claim involves edges of the original graph G, not H.

Proof. Let $C = V(G) \setminus B$, noting that $|C|^\alpha \geq (\frac{1}{2}n)^\alpha \geq \frac{1}{2}n^\alpha$, and since $d \geq \frac{1}{2}\gamma n^\alpha$ by Proposition 30.1, we conclude

$$e_G(B, T) \leq \frac{1}{8}\alpha\gamma|B||C|^\alpha.$$

Note that

$$E(G) = E(G[B \cup S]) \cup E(G[C]) \cup E(G[B, T]),$$

where $E(G[B', T])$ denotes the set of edges of G with one end in B' and the other in T. To see this, we note that vertices of B can only be adjacent to vertices of $B \cup N_G(B) = B \cup S \cup T$. With this we have

$$\begin{aligned} e_G(B \cup S) &\geq e(G) - e(G[C]) - e_G(B, T) \\ &\geq \frac{1}{2}\gamma(|B| + |C|)^{1+\alpha} - \frac{1}{2}\gamma|C|^{1+\alpha} - \frac{1}{8}\alpha\gamma|B||C|^\alpha, \end{aligned}$$

where this inequality used $e(G) = \frac{1}{2}\gamma n^{1+\alpha} = \frac{1}{2}\gamma(|B'| + |C|)^{1+\alpha}$, α-maximality, and the inequality noted above. Note that

$$(|B| + |C|)^{1+\alpha} = |C|^{1+\alpha}(1 + |B|/|C|)^{1+\alpha} \geq |C|^{1+\alpha} + (1 + \alpha)|B||C|^\alpha.$$

Using this gives

$$e_G(B \cup S) \geq \frac{1}{2}\gamma(1 + \alpha)|B|C|^\alpha - \frac{1}{8}\alpha\gamma|B||C|^\alpha \geq \frac{1}{2}\gamma(1 + \frac{1}{2}\alpha)|B||C|^\alpha.$$

By α-maximality we have $e_G(B \cup S) \leq \frac{1}{2}\gamma(|B| + |S|)^{1+\alpha}$, so in total this implies

$$|S| \geq \left((1 + \frac{1}{2}\alpha)|B||C|^\alpha\right)^{1/(1+\alpha)} - |B|.$$

For $w \in S$, let X_w be the indicator random variable for the event $w \notin N_{H_Q}(B)$. Then

$$\Pr[X_w = 1] = (1 - p_c)^{d_{H_Q}(w)} \leq (1 - p_c)^{\Delta} \leq e^{-\lambda^{1/2}}.$$

Thus if $X = \sum_{w \in S} X_w$ then $\mathbb{E}[X] \leq |S|e^{-\lambda^{1/2}}$, and by Markov's inequality this means $\Pr[X \geq |S|/2] \leq 2e^{-\lambda^{1/2}}$. This gives the result. $\square$

From now on we assume $e_G(B,T) > d\alpha|B|/16$. For each $w \in T$, let Y_w be the indicator random variable for the event $w \in N_{H_Q}(B)$. Then by the inequality $(1-a)^b \geq 1 - \frac{1}{2}ab$ for $ab < \frac{1}{2}$,

$$\mathbb{E}[Y_w] = 1 - (1 - p_c)^{d_H(w)} \geq \frac{1}{2} \min\{1, p_c \cdot d_H(w)\}.$$

We partition T into two sets based on which term prevails in this minimum. Namely, let $T_1 = \{w \in T : d_H(w) \leq p_c^{-1}\}$ and $T_2 = T \setminus T_1$. If $Y = \sum_{w \in T} Y_w$, then

$$\mathbb{E}[Y] \geq \sum_{w \in t_1} \frac{1}{2} p_c \cdot d_H(w) + \frac{1}{2}|T_2| = \frac{1}{2} p_c e_H(B, T_1) + \frac{1}{2}|T_2| \geq \frac{1}{2} p_c e_H(B, T_1) + \frac{1}{2}\lambda^{-1/2} p_c e_H(B, T_2),$$

where this last step used that each vertex of $T_2 \subseteq T$ has degree at most $\Delta = \lambda^{1/2} p_c^{-1}$ in H. Thus

$$\mathbb{E}[Y] \geq \frac{1}{2}\lambda^{-1/2} p_c e_H(B, T).$$

By hypothesis,

$$e_H(B,T) \geq e_G(B,T) - \sum_{v \in B} |\phi(v)| \geq \frac{1}{32} d\alpha |B|.$$

This together with the hypothesis $d \geq \lambda^{1/2}(p_c \cdot \alpha)^{-1}$ implies $\mathbb{E}[Y] \geq \frac{1}{64}\lambda^{1/2}|B|$.

Note that Y is a function of which colors survive in Q. Each color appears at most $|B|$ times since G is properly colored, so changing Q by a single element changes Y by at most $|B|$, i.e. Y is $|B|$-Lipschitz. By the multiplicative Azuma inequality (Lemma 4.8), we have $\Pr[Y \leq \frac{1}{2}\mathbb{E}[Y]] \leq e^{-\Omega(\lambda)}$. Since $Y = |N_{H_Q}(B)|$, we conclude the result.

$\square$

Proof of Lemma 30.3. Similar to our proofs involving spread hypergraphs, we will iteratively generate random sets Q_i a total of $\ell = 100\alpha^{-1} \log\log(n)$ times and take $Q = \bigcup Q_i$, iteratively arguing that each Q_i is likely to have good properties.

Let q_c be the unique solution to $p_c = 1 - (1 - q_c)^{\ell}$; the main take away being that $q_c = \Omega(p_c/\ell)$. For $1 \leq i \leq \ell$, let Q_i be obtained by including each color of R independently and with probability q_c (and independent of any other Q_j set), noting that $\bigcup_{i=1}^{\ell} Q_i$ has the same distribution as Q.

Let B_i be the set of vertices x that can be reached from v by some $(Q_1 \cup \cdots \cup Q_i)$-rainbow path P_x of length at most i. Let $\phi_i : V(G) \to 2^{V(G) \cup R}$ be the function which maps to the vertices and colors of P_x if $x \in B_i$, and otherwise $\phi(x) = \emptyset$. Note that $|\phi_i(x)| \leq 2i \leq 2\ell$. We wish to show that $|B_i|$ is rapidly increasing with high probability.

First note that B_1 is just the set of neighbors of v with $c(vx) \in Q_1$. By Proposition 30.1 we have $d(v) \geq \frac{1}{2}d$, so $\mathbb{E}[|B_1|] \geq \frac{1}{2}dq_c$. Thus by the Chernoff bound we will have $|B_1| \geq \frac{1}{4}dq_c > 2\lambda^2 q_c^{-1}$ with high probability, so from now on we assume this is the case..

Note that $N_{Q_i,\phi_i}(B_i \cap U_i) \subseteq B_{i+1}$. Since $d \geq \lambda^2(\alpha \cdot q_c)^{-1}n^\alpha$ by hypothesis, and since $|B_i| \geq |B_1| > 2\lambda^2 q_c^{-1}$, as long as $|B_i| < n/3$ we can apply Lemma 30.4 to get

$$N_{Q_i,\phi_i}(B_i) \geq \frac{1}{4}|B_i| \min\left\{\frac{d \cdot q_c \cdot \alpha}{64\lambda^{1/2}}, \left(\frac{n}{2|B_i|}\right)^{\alpha/(1+\alpha)} - 1\right\}$$

with probability at least $1 - e^{-\Omega(\lambda^{1/2})}$. Note that the leftside of the minimum is always at least n^α, so the minimum is always achieved by the righthand side. Using this and $\alpha \leq 1$ gives

$$|N_{Q_i,\phi_i}(B_i)| \geq \frac{1}{4}|B_i|((n/2|B_i|)^{\alpha/2)} - 1)$$

with probability at least $1 - e^{-\Omega(\lambda^{1/2})}$. Thus with probability at least $1 - \ell e^{-\Omega(\lambda^{1/2})} = 1 - O(\alpha^{-1}e^{-\Omega(\lambda^{1/2})})$ this holds for all i (note that the $\log\log(n)$ gets absorbed by $e^{-\Omega(\lambda^{1/2})}$) since $\lambda \geq (\log\log n)^2)$. We claim that this implies $|B_{\ell-1}| \geq n/3$. And indeed, using that $|B_{i+1}| \geq |B_i| + |N_{Q_i,\phi_i}(B_i)|$ (since B_i and $N(B_i) \supseteq N_{Q_i,\phi_i}(B_i)$ by definition), one can prove by induction that $|B_i| \geq (n/2)^{1-(1-\alpha/16)^i}$ provided $|B_{i-1}| \leq n/3$. This gives the result. $\square$

Tomon [142] proved several other nice results using a result which extends Lemma 30.3 in two ways. The first way is by enforcing short paths from v provided we don't require v to reach as many vertices (and it is easy to adapt our current proof to achieve this end). The other extension is that it allows one to sample a random set of vertices $U \subseteq V(G)$ in addition to a random set of colors, and which guarantees short paths from v to a large set of vertices. To state such a result, we say that a path is a (U, Q)-rainbow path if it is a rainbow path whose internal vertices all lie in U and whose colors all lie in U.

Lemma 30.6. *There exists a sufficiently large constant C such that the following holds. Let $p, p_c, \alpha \in (0,1]$, n a positive integer, $\tau \in [1/\log_3 n, \frac{1}{2})$, and $\lambda > C(\log\log n)^1 0$. Let G be an n-vertex α-maximal graph with proper edge coloring $c : E(G) \to R$ with average degree $d = d(G)$ satisfying either $d > C\lambda^2(\alpha^2 \cdot p_c^2)^{-1}n^\alpha$ if $p = 1$, and otherwise $d > C\lambda^2(\alpha^3 \cdot p \cdot p_c^2)^{-1}n^\alpha$.*

Let $U \subseteq V(G)$ be obtained by including each vertex independently and with probability p, and similarly define $Q \subseteq R$. For each $v \in V(G)$, with probability at least $1 - O(\alpha^{-1}e^{-\Omega(\lambda^{1/2})})$, at least $n^{1-\tau}$ vertices of G can be reached from v by a (U, Q)-path of length at most $O(\alpha^{-1}\log(1/\tau))$.

To prove this, one needs to extend Lemma 30.4 to say that the same conclusion holds for $N_{Q,\phi}(U)$ with $U \subseteq B$ obtained by including each vertex independently and with probability p. This isn't too hard to prove if the vertices of B all have reasonable maximum degree, and one extra case deals with the situation where this doesn't happen.

With Lemma 30.6 it is possible to prove results about rainbow Turán numbers of subdivisions of K_t. To this end, let $\mathcal{K}_t$ denote the set of subdivisions of K_t (i.e. the graphs which can be obtained by subdividing each edge of K_t some number of times). Mader [111] showed $\mathrm{ex}(n, \mathcal{K}_t) = O_t(n)$, and again the hypercube shows $\mathrm{ex}^*(n, \mathcal{K}_t) = \Omega(n \log n)$ for $t \geq 3$. Jiang, Letzter, Methuku, and Yepremyan [93] showed $\mathrm{ex}^*(n, \mathcal{K}_t) = O((\log n)^{60} n)$. These bounds were improved significantly by Tomon [142].

Theorem 30.7 ([142]). *For all fixed t we have* $\mathrm{ex}^*(n, \mathcal{K}_t) \le (\log n)^{6+o(1)} n$.

Proof Sketch. Let G be an n-vertex graph with $e(G) \ge 2(\log n)^{6+\epsilon} n$ and $c : E(G) \to R$ a proper coloring, and let $\alpha = 1/\log_2(G),\ s = (\log n)^{1+\epsilon/10},\ p = p_c = 1/s$, and $\lambda = (\log n)^{\epsilon/10}$. For some slight ease of notation we assume G is α-maximal (though it's easy for the rest of the proof to go through by reducing to an α-maximal subgraph). Note that by our choice of parameters, $d(G) \ge C\lambda(\alpha^3 p \cdot p_c^2)^{-1} n^{\alpha}$.

Define an auxiliary graph H where two vertices v, w are adjacent if there exist at least $s/6$ internally disjoint paths from v to w such that no color is used more than once int he union of the paths.

We claim that if H has minimum degree at least $n/6$, then G contains a rainbow K_t-subdivision. Indeed, by Theorem 5.4, H (easily) contains a 1-subdivision of K_t. One can then greedily replace each edge with a rainbow path which doesn't use any vertices or colors that have already been used.

To show that H has this minimum degree, we apply Lemma 30.6 with the stated parameters and $\tau = 1/\log_3(n)$ to any vertex v. By a similar argument to before, this implies there exist partitions $U_1, \ldots, U_s$ and $Q_1, \ldots, Q_s$ such that the sets B_i of vertices we can reach from v with a (U_i, Q_i)-rainbow path all have size at least $n/3$. This implies that there exist at least $n/6$ vertices w in at least s of the B_i sets, proving that $d_H(v) \ge n/6$. $\square$

31 The Linear Programming Method

Throughout this section we let bold letters $\mathbf{x}$ denote vectors and write x_i to denote the ith coordinate of $\mathbf{x}$. For vectors $\mathbf{x}, \mathbf{y}$ of the same dimension, we write $\mathbf{x} \le \mathbf{y}$ to mean $x_i \le y_i$ for all i. Our writing throughout will be somewhat informal in order to convey the main ideas of the method without getting too bogged down in technicalities. More precise writing can be found in the book by Matoušek and Gärtner [115] dedicated to this topic.

31.1 Linear Programming Basics

Many problems in extremal combinatorics can be phrased in terms of *integer programs*, i.e. as problems which aim to either maximize or minimize a linear function $\sum c_i x_i$ (called the *objective function*) where each x_i is an integer-valued variable satisfying a set of linear inequalities and equalities. An assignment of each of the x_i variables to some value such that they satisfy these constraints is called a *feasible solution* to the program, and the maximum value of the objective function that is achieved by a feasible solution is called the *optimal value* or *optimum value* of the program.

As a very basic example, consider the integer program (I) which has variables $x_1, x_2 \in \mathbb{Z}$ defined by

$$\begin{aligned} \text{maximize } & x_1 + x_2 \\ \text{subject to } & x_1 \le 1.5, \\ & x_2 \le 1.5. \end{aligned}$$

In this example, $x_1 + x_2$ is the objective function, $x_1, x_2 \le 1.5$ are our only constraints, and $x_1 = 1$, $x_2 = 1$ is a feasible solution. This feasible solution shows the optimal value of the program is at least 2, and it is not hard to see that it is exactly 2 since x_1, x_2 must be integers.

For a more interesting example, consider the integer program (M) which has variables $x_{i,j} \in \mathbb{Z}$ for all distinct $i, j \in [n]$ and is defined by

$$\begin{aligned} \text{maximize } & \sum_{1 \le i < j \le n} x_{i,j} \\ \text{subject to } & 0 \le x_{i,j} \le 1 \quad \forall i \ne j \\ & x_{i,j} + x_{i,k} + x_{j,k} \le 2 \quad \forall i \ne j \ne k. \end{aligned}$$

This program seems arbitrary at first, until one realizes its connection to a classical extremal combinatorics problem.

Lemma 31.1. *The optimal value of (M) equals* $\mathrm{ex}(n, K_3)$.

Proof. Given a triangle-free graph G on $[n]$ with $\mathrm{ex}(n, K_3)$ edges, define $x_{i,j}$ by setting $x_{i,j} = 1$ if $ij \in G$ and $x_{i,j} = 0$ otherwise. It is not difficult to see that this is a feasible solution to (M), and in particular the constraint $x_{i,j} + x_{i,k} + x_{j,k} \le 2$ is satisfied because G is triangle-free. The optimal value of (M) is thus at least $\sum x_{i,j} = e(G) = \mathrm{ex}(n, K_3)$.

On the other hand, say we have some assignment $x_{i,j}$ to (M) achieving its optimal value. Observe that $x_{i,j} \in \{0,1\}$ for all i,j since $0 \leq x_{i,j} \leq 1$ must be an integer. In this case we define a graph G on $[n]$ by setting $ij \in G$ if and only if $x_{i,j} = 1$, and reversing the logic above shows that G is triangle-free with at least as many edges as the optimal value of (M), proving the result. $\square$

The observation above shows that Mantel's Theorem is equivalent to saying that the optimal value of the integer program (M) is $\lfloor n^2/4 \rfloor$. Using similar ideas, one can rephrase many problems in extremal combinatorics in terms of finding the solution to some integer program (such as computing $\mathrm{ex}(n,F)$ for any graph F, for example). Given this, the question now becomes: how do we find the optimal value of an integer program in general?

As in our first example, if we have an integer program (I) which aims to maximize some objective function ϕ, then obtaining lower bounds is pretty easy: we just plug into ϕ some explicit feasible solution $\mathbf{x}$, at which point $\phi(\mathbf{x})$ becomes a valid lower bound for the optimal value. On the other hand, proving upper bounds is quite difficult in general. Because of this, we will try and make things easier for ourselves by removing the requirement that the variables x_i be integers, and instead allow them to be any real number. We will call such a relaxed problem a *linear program*, and it will turn out that these programs are substantially easier to solve.

For example, the linear relaxation of our very first example (I) gives the linear program (L) which has variables $x_1, x_2 \in \mathbb{R}$ and is defined by

$$\begin{gathered}\text{maximize } x_1 + x_2 \\ \text{subject to } x_1, x_2 \leq 1.5.\end{gathered}$$

Because this is a *linear* program and not an *integer* program, $x_1 = x_2 = 1.5$ now becomes a feasible solution for (L) and achieves the optimal value of 3 for this program.

In the setting of linear programs we again must ask: how do we determine their optimal values in general? If our program aims to maximize an objective function, then as before lower bounds simply require us to plug in an explicit set of values into the variables of our objective function. Crucially (and non-obviously), it turns out upper bounds for linear programs can be achieved in the *exact same way.*

Somewhat more precisely, it turns out that for every linear program (L) which tries to *maximize* some objective function ϕ, there exists a corresponding "dual program" (D) which is another linear program that aims to *minimize* an objective function ψ where, crucially, the optimal minimal value of (D) is equal to the optimal maximum value of (L). This gives the following framework for solving problems using linear programming:

1. Phrase your extremal problem in terms of an integer program (I) which tries to, say, maximize some objective function ϕ.

2. Consider the linear relaxation (L) of (I) and then take its corresponding dual program (D) which aims to minimize some objective function ψ.

3. Find an explicit feasible solution for (D) and plug this into ψ. This value we obtain is an upper bound for the optimal value of (D) (since (D) aims to minimize some function),

which is also an upper bound for the optimal value of (L), which is an upper bound for the optimal value of our original program (I) corresponding to our given extremal problem.

While this approach can be used to give effective upper bounds for a wide variety of problems, we must warn the reader that these bounds can sometimes be quite far from tight. For example, in the triangle-free integer program (M) mentioned above, if we take its linear relaxation (M'), then one can check that setting $x_{i,j} = 2/3$ for all i, j is a feasible solution, which shows the optimum value of this linear relaxation of (M') is at least $\frac{2}{3}\binom{n}{2}$. As this is is much larger than the true optimal value of $\lfloor n^2/4 \rfloor$ for (M), we see that this linear programming approach is not particularly useful for the problem of upper bounding $\text{ex}(n, K_3)$. That being said, there are many cases where the gap between the optimal value of an integer program and its corresponding linear program is very small (sometimes even 0), in which case this approach can give relatively simple proofs for fairly difficult results.

31.2 Weak Duality

Our discussion above about upper bounding linear programs was rather vague, so let us consider a concrete example of what we are talking about. Perhaps the simplest non-trivial linear program is the following:

$$\begin{aligned} &\text{maximize } x_1 + x_2 \\ &\text{subject to } x_1, x_2 \leq 1. \end{aligned}$$

It is (hopefully) clear that this linear program has optimal value 2, but let us pause for a second and formally show why this is the case.

First off, we have a lower bound of 2 for the optimal value simply because $x_1 = x_2 = 1$ is a feasible solution which achieves the value of 2 in the objective function. For the upper bound, we observe that adding the inequality $x_1 \leq 1$ to the inequality $x_2 \leq 1$ gives the new inequality $x_1 + x_2 \leq 2$, which is exactly the upper bound on the objective function we wanted to show.

One can generalize the idea above by considering arbitrary non-negative linear combination of the constraints, which gives the following result. Here and throughout, we encode the constraints of our linear program as an inequality involving vectors and matrices: specifically, we write $A\mathbf{x} \leq \mathbf{b}$ to encode the set of inequalities $\sum_j A_{i,j} x_j \leq b_i$ for all i.

Lemma 31.2. *Let A be an $m \times n$ matrix and $\mathbf{b}, \mathbf{c}$ vectors of dimensions m, n respectively. If (P) is the linear program with variables $\mathbf{x} \in \mathbb{R}^m$ defined by*

$$\begin{aligned} &\text{maximize } \mathbf{c}^T\mathbf{x} \\ &\text{subject to } A\mathbf{x} \leq \mathbf{b}, \end{aligned}$$

and if there exists a non-negative vector $\mathbf{y}$ of dimension m such that $\mathbf{y}^T A = \mathbf{c}^T$, then (P) has optimum value at most $\mathbf{y}^T\mathbf{b}$.

For instance, our example above corresponds to $A = I_2$ and $\mathbf{b} = \mathbf{c} = \mathbf{y} = (1, 1)$, giving the desired upper bound of 2 in this case.

Proof. By assumption we have the constraints $\sum_j A_{i,j}x_j \leq b_i$ for all i. Because each y_i is non-negative by assumption, we also have the constraints $\sum y_i A_{i,j}x_j \leq y_i b_i$ for all i. Adding all of these new constraints together gives

$$\sum_{i,j} y_i A_{i,j} x_j \leq \sum_i y_i b_i.$$

Observe that the lefthand side equals $\mathbf{y}^T A\mathbf{x}$, which in turn equals $\mathbf{c}^T\mathbf{x}$ by hypothesis. Similarly the righthand side equals $\mathbf{y}^T\mathbf{b}$, so the inequality above is exactly the result we aimed to prove. □

In general there may be many choices of $\mathbf{y}$ satisfying the hypothesis above, and we can get the best possible upper bound for (P) by choosing the $\mathbf{y}$ which minimizes $\mathbf{y}^T\mathbf{b}$. This idea quickly gives the following.

Corollary 31.3 (Weak Duality Theorem I)**.** *Let A be an $m \times n$ matrix and $\mathbf{b}, \mathbf{c}$ vectors of dimensions m, n respectively. Let (P) denote the linear program with variables $\mathbf{x} \in \mathbb{R}^m$ defined by*

$$\begin{aligned} &\text{maximize } \mathbf{c}^T\mathbf{x} \\ &\text{subject to } A\mathbf{x} \leq \mathbf{b}. \end{aligned}$$

Let (D) denote the linear program with variables $\mathbf{y} \in \mathbb{R}^n$ defined by

$$\begin{aligned} &\text{minimize } \mathbf{b}^T\mathbf{y} \\ &\text{subject to } A^T\mathbf{y} = \mathbf{c} \\ &\qquad\qquad \mathbf{y} \geq 0. \end{aligned}$$

Then the optimum value of (D) is an upper bound for the optimum value of (P).

That is, for every linear program in the form of (P), which is typically called the "primal" program, there exists an (easy to construct) linear program (D) called the "dual" program whose optimum value upper bounds the optimal value of (P). Moreover, since the objective function of (D) aims to be minimized, it is significantly easier to upper bound its optimum value compared to (P) (since to upper bound (D) we just need to plug in some feasible solution into its objective function).

The equality constraint $A^T\mathbf{y} = \mathbf{c}$ in the Weak Duality Theorem can be a little restrictive to work with. One can weaken this condition into an inequality at the cost of requiring the variables $\mathbf{x}$ to be non-negative (which often holds in practice). Doing this gives the following (ultimately equivalent) version of weak duality which is more symmetric in (P) and (D).

Theorem 31.4 (Weak Duality Theorem II)**.** *Let A be an $m \times n$ matrix and $\mathbf{b}, \mathbf{c}$ vectors of dimensions m, n respectively. Let (P) denote the linear program with variables $\mathbf{x} \in \mathbb{R}^m$ defined by*

$$\begin{aligned} &\text{maximize } \mathbf{c}^T\mathbf{x} \\ &\text{subject to } A\mathbf{x} \leq \mathbf{b} \\ &\qquad\qquad \mathbf{x} \geq 0, \end{aligned}$$

and let (D) denote the linear program with variables $\mathbf{y} \in \mathbb{R}^n$ *defined by*

$$\begin{aligned} \text{minimize } & \mathbf{b}^T\mathbf{y} \\ \text{subject to } & A^T\mathbf{y} \geq \mathbf{c} \\ & \mathbf{y} \geq 0. \end{aligned}$$

Then the optimum value of (D) is an upper bound for the optimum value of (P).

Sketch of Proof. Given any feasible solution $\mathbf{y}$ to (D), we take linear combinations of the constraints $A\mathbf{x} \leq \mathbf{b}$ as in Lemma 31.2 to conclude that for any feasible solution $\mathbf{x}$ to (P),

$$\mathbf{y}^T\mathbf{b} \geq \mathbf{y}^T A\mathbf{x} \geq \mathbf{c}^T\mathbf{x},$$

where this last inequality used $A^T\mathbf{y} \geq \mathbf{c}$ and $\mathbf{x} \geq 0$. Taking a minimum over all such $\mathbf{y}$ gives the result. □

In many applications, the most natural linear program we'd like to use won't be of the exact form as stated in Theorem 31.4. However, there exist a number of standard tricks that can be used to reduce every linear program into the "standard" form written out in Theorem 31.4.

As a simple example, if we have a linear program (P) where we want to *minimize* some function ϕ instead of maximizing it, then we can simply consider the linear program (P') defined in the same way as (P) except that it aims to *maximize* the function $-\phi$. With this, the optimal value of (P') is just the negation of that of (P), so it suffices to work only with (P').

As another example, say we are considering some linear program (P) which meets the hypothesis of Theorem 31.4 except that it has one variable x which we do not want to assume is non-negative. In this case, we can create a new program (P') obtained by adding two new variables y, z, replacing every instance of x in our objective function and constraints with the expression $y - z$, and then adding the non-negativity constraints $y, z \geq 0$. It is not difficult to see[1] that this new program (P') has the same optimum value of (P).

While the ideas mentioned above can be used to translate any linear program into the form of Theorem 31.4, in practice it is often simpler to consider more general versions of Theorem 31.4 which allow for more diverse sets of constraints. In particular, an easy adaptation of the proof of Theorem 31.4 yields the following "Dualization Recipe" which is copied verbatim form Matoušek and Gärtner [115] and which illustrates how to efficiently translates any primal program (P) that aims to maximize some function into a corresponding dual program. Programs which aim to minimize a function can be translated in the same way after either replacing their objective function with its negation or their variables with their negations. While either of these two fixes will fix the maximization problem, they may give different looking dual programs, and one should play around with which one gives the cleaner problem to work with.

[1] Namely, if an optimal solution in (P) has $x \geq 0$, then we get this same optimum value in (P') by taking $y = x$ and $z = 0$, and if $x \leq 0$ we can take $y = 0$ and $z = -x$

Dualization Recipe

	Primal linear program	Dual linear program
Variables	$x_1, x_2, \ldots, x_n$	$y_1, y_2, \ldots, y_m$
Matrix	A	A^T
Right-hand side	$\mathbf{b}$	$\mathbf{c}$
Objective function	$\max \mathbf{c}^T\mathbf{x}$	$\min \mathbf{b}^T\mathbf{y}$
Constraints	ith constraint has $\leq$ $\geq$ $=$	$y_i \geq 0$ $y_i \leq 0$ $y_i \in \mathbb{R}$
	$x_j \geq 0$ $x_j \leq 0$ $x_j \in \mathbb{R}$	jth constraint has $\geq$ $\leq$ $=$

.

31.3 Applications to Extremal Combinatorics

Here we look at two examples of using linear programs for extremal combiantorics. For both of these applications, we will make use of our Weak Duality Theorem II after massaging our initial linear programs into the appropriate form needed to apply this theorem. Ultimately this approach will yield dual programs that are not as elegant as one would get by using some form of the Dualization Recipe mentioned above, but we persist in using this crude approach in order to emphasize the "mechanical" nature of how to use Weak Duality in practice.

31.3.1 Covering Grids

Our first example concerns covering grids with hyperplanes. The classic result in this area is the Alon-Füredi Theorem 26.3, which we recall says that if $h_1, \ldots, h_m$ are a sequence of hyperplanes in $\mathbb{R}^d$ which do not contain the origin and whose union contains every other point of $\{0,1\}^d$, then $m \geq d$. Here we look at a two-way generalization of this problem as follows.

Definition 6. Given an integer $k \geq 1$ and sets $S_1, \ldots, S_d \subseteq \mathbb{R}^d$ all containing 0, we say that a sequence of hyperplanes $h_1, \ldots, h_m$ is a *k-punctured cover* if no hyperplane h_i contains the origin $\vec{0}$ and if every other point $p \in \prod S_i \setminus \{\vec{0}\}$ is contained in at least k hyperplanes. We let $\mathrm{cov}_k(S_1, \ldots, S_d)$ denote the smallest value m such that there exists a k-punctured cover $h_1, \ldots, h_m$ for $S_1, \ldots, S_d$.

For example, if $k = 1$ and $S_i = \{0,1\}$ for all i then the Alon-Füredi Theorem exactly says $\mathrm{cov}_1(S_1, \ldots, S_n) = d$. A straightforward trivial construction gives the general upper bound

$\mathrm{cov}_k(S_1, \ldots, S_d) \le k \sum_i (|S_i| - 1)$. The best known general lower bound uses a variant of the Combinatorial Nullstellensatz due to Ball and Serra [11] and gives

$$\mathrm{cov}_k(S_1, \ldots, S_d) \ge \sum_i (|S_i| - 1) + (k-1) \max_i (|S_i| - 1).$$

Both of these bounds are obviously tight when $d = 1$. The $d = 2$ case was systematically studied by Bishnoi, Boyadzhiyska, Das, and den Bakker [23], who showed that the lower bound of Ball and Serra is tight provided $|S_1| \ge k|S_2|$. However, by adapting a linear programming approach due to Clifton and Huang [38], they were able to greatly improve upon the Ball and Serra bound whenever $|S_1| \approx |S_2|$. For simplicity we will only prove the following, where here we say that a point $p \in S_1 \times S_2$ is an *axis point* if it lies on either the x-axis or y-axis, and we say that p is a *generic point* otherwise.

Theorem 31.5 ([23])**.** *If $S_1, S_2 \subseteq \mathbb{R}$ are such that $0 \in S_1 \cap S_2$ and $|S_1| = |S_2| = n$, then $\mathrm{cov}_k(S_1, S_2) \le \lceil 3k/2 \rceil (n-1)$. Moreover, if every line in $\mathbb{R}^2$ which contains at least two axis points of $S_1 \times S_2$ contains no generic points, then*

$$\frac{3}{2}k(n-1) \le \mathrm{cov}_k(S_1, S_2) \le \left\lceil \frac{3}{2}k \right\rceil (n-1).$$

Note that if S_1, S_2 are "random" sets of size n then this moreover condition applies with probability 1, so this result solves the problem for "most" grids $S_1 \times S_2$ of equal side lengths.

Proof. The result is trivial if $n = 1$, so we assume $n \ge 2$ from now on. We omit the proof of the upper which follows by constructing some (not too complicated) explicit cover, and we instead focus on the lower bound which uses the linear programming method.

Let's begin by informally talking out loud about how to derive an appropriate linear program to work with. The objective function should ultimately measure how many lines (i.e. hyperplanes) we use in our k-punctured cover. To this end, for each line ℓ we introduce a variable x_ℓ which measures how many copies of ℓ we use in our cover and then take $\sum_\ell x_\ell$ to be our objective function. There is a snag in this approach, which is that naively we are summing over an infinite set of lines ℓ, making our sum ill-defined. However, it is easy to reduce the sum down to a finite set: take $\mathcal{L}$ to be the set of lines which contain at least two points of $S_1 \times S_2$ and which do not contain $(0,0)$, noting that this is a finite set. It is not difficult to see that there exists a k-punctured cover of $S_1 \times S_2$ of size $\mathrm{cov}_k(S_1, S_2)$ which only uses lines from $\mathcal{L}$, since any line $\ell \notin \mathcal{L}$ (which covers at most 1 point $(a,b) \ne (0,0)$ of $S_1 \times S_2$ by definition) used in the cover can be replaced by some $\ell' \in \mathcal{L}$ which also covers (a,b) while maintaining that this is a k-punctured cover.

With the above in mind, we define a linear program (P) with variables x_ℓ for each $\ell \in \mathcal{L}$ by

$$\begin{aligned} \text{minimize } & \sum_{\ell \in \mathcal{L}} x_\ell \\ \text{subject to } & \sum_{\ell \ni p} x_\ell \ge k \quad \forall p \in S_1 \times S_2 \setminus \{(0,0)\} \\ & \mathbf{x} \ge 0. \end{aligned}$$

Observe that if (P) were an integer program then its optimal value would exactly equal $\mathrm{cov}_k(S_1, S_2)$, and as a linear program its optimal value (which can only be smaller than that of its corresponding integer program) serves as a lower bound for $\mathrm{cov}_k(S_1, S_2)$. In total this gives the following.

Claim 31.6. *To prove the result, it suffices to show the optimal value of (P) is at least* $\frac{3}{2}k(n-1)$.

At this point a real expert in using the linear programming method would immediately take the appropriate dual of (P) and go from there, but we'll take things a little slowly here. In particular, we will first replace (P) with an equivalent linear program (P') which is in the format needed to apply our (stated version of) the Weak Duality Theorem, giving the following equivalent version of the previous claim.

Claim 31.7. *To prove the result, it suffices to show the optimal value of the linear program (P') with variables* x_ℓ *for* $\ell \in \mathcal{L}$ *defined by*

$$\begin{aligned} \text{maximize } & \sum_{\ell \in \mathcal{L}} -x_\ell \\ \text{subject to } & \sum_{\ell \ni p} -x_\ell \le -k \quad \forall p \in S_1 \times S_2 \setminus \{(0,0)\} \\ & \mathbf{x} \ge 0 \end{aligned}$$

has optimum value at most $-\frac{3}{2}k(n-1)$.

In the language of Theorem 31.4, we have $c = (-1, -1, \ldots)$, $b = (-k, -k, \ldots)$, and A the matrix with rows indexed by $S_1 \times S_2 \setminus \{(0,0)\}$ and columns indexed by $\mathcal{L}$ such that $A_{p,\ell} = -1$ if $p \in \ell$ and $A_{p,\ell} = 0$ otherwise. Applying Theorem 31.4 then gives the following.

Claim 31.8. *To prove the result, it suffices to show the optimal value of the linear program (D') with variables* y_p *for* $p \in S_1 \times S_2 \setminus \{(0,0)\}$ *defined by*

$$\begin{aligned} \text{minimize } & \sum_{p \in S_1 \times S_2 \setminus \{(0,0)\}} -k y_p \\ \text{subject to } & \sum_{p \in \ell} -y_\ell \ge -1 \quad \forall \ell \in \mathcal{L} \\ & \mathbf{y} \ge 0 \end{aligned}$$

has optimum value at most $-\frac{3}{2}k(n-1)$.

All that remains now is to choose some specific weightings for the y_p which satisfy the constraints above and which have $\sum y_p \ge \frac{3}{2}k(n-1)$ sufficiently large. Since our hypothesis on S_1, S_2 mirrors the case that these sets are random, there is not much structure for us to work with in determining our weighting other than the fact that some points are axis-points while others are generic. Because of this, we might first try the very naive strategy of choosing some $\alpha, \gamma \ge 0$ and assigning $y_p = \alpha$ for every axis-point p and $y_p = \gamma$ for every generic-point. It remains to check which (smallest) values of α, γ make this a feasible solution for (D').

To this end, the only thing that we need to check is that every line $\ell \in \mathcal{L}$ has $\sum y_p \le 1$. By our hypothesis, any line $\ell \in \mathcal{L}$ which contains at least two axis-points contains only these two axis-points (note that $\ell \in \mathcal{L}$ can not be equal to either the x-axis or the y-axis because it must avoid $(0,0)$ by definition of $\mathcal{L}$), so all of these lines give the constraint

$$2\alpha \leq 1.$$

Otherwise, ℓ can contain at most one axis-point and at most $n-1$ generic-points, giving the constraint

$$\alpha + (n-1)\gamma \leq 1.$$

Because there are $(2n-2)$ axis-points and $(n-1)^2$ generic points, these conditions imply that our objective function satisfies

$$-k\sum_p y_p = -k((2n-2)\alpha + (n-1)^2\gamma) = -k(n-1)\cdot(\alpha+\alpha+(n-1)\gamma) \leq -k(n-1)\cdot(1/2+1)\,.$$

Moreover, equality holds only if $\alpha = 1/2$ and $\gamma = \frac{1}{2(n-1)}$. One can check that taking these values for α, γ does indeed give a feasible solution to (D') with the desired value for its objective function, proving the result. $\square$

Many more results in the spirit of Theorem 31.5 are proven in [23]. For example, an essentially identical proof gives the same asymptotic lower bound if we weaken the hypothesis of Theorem 31.5 to allow a bounded number of generic-points on lines containing two axis-points. A very similar argument can also be used to give essentially tight lower bounds even if we drop the hypothesis $|S_1| = |S_2|$, though in this case one needs to use two different weights for axis-points depending on if they lie on the shorter or the longer axis.

31.3.2 Average Sizes in Antichains

Our next result concerns antichains.

Theorem 31.9. *If $\mathcal{F} \subseteq 2^{[n]}$ is an antichain with $|\mathcal{F}| \geq \binom{n}{r}$ for some $r \leq n/2$, then*

$$|\mathcal{F}|^{-1}\sum_{A\in\mathcal{F}} |A| \geq r.$$

Equivalently, this says that if $|\mathcal{F}| \geq \binom{n}{r}$, then the average size of its elements is at least r, which is best possible by considering $\mathcal{F} = \binom{[n]}{r}$. This result was original proven by Kleitman and Milner [104] in two ways, and also proven independently by Hochberg [89] with another approach. Here we present the linear programming proof of Kleitman and Milner.

Proof. We will prove this result using the linear programming method. We note, however, that it is not at all obvious that this should be a reasonable approach to try for this problem[1]. Indeed, there does not seem to be any way of encoding the fact that $\mathcal{F}$ is an antichain through linear constraints. The key insight is that, although there is no linear constraints that hold

[1] Other than the fact that we have put this result in a chapter about using linear programming, of course

iff $\mathcal{F}$ is an antichain, there is an important linear constraint which every antichain satisfies, namely the LYM inequality Corollary 2.5, which states that

$$\sum_{A\in\mathcal{F}}\binom{n}{|A|}^{-1}\le 1.$$

From this we can conclude the following.

Claim 31.10. *To prove the result, it suffices to show that the linear program (P) with variables $\mathbf{x}\in\mathbb{R}^n$ defined by*

$$\begin{aligned}\text{minimize } & \sum_i ix_i\\ \text{subject to } & \sum_i x_i\ge\binom{n}{r}\\ & \sum_i\binom{n}{i}^{-1}x_i\le 1\\ & \mathbf{x}\ge 0\end{aligned}$$

has optimum value at least $r\binom{n}{r}$.

Proof. Assume that we could do this and let $\mathcal{F}$ be an antichain of size at least $\binom{n}{r}$. Let x_i denote the number of elements of $\mathcal{F}$ of size i, noting that this implies $x_i\ge 0$, that $\sum x_i\ge\binom{n}{r}$, and that $\sum\binom{n}{i}^{-1}x_i\le 1$ by the LYM inequality. It follows from our hypothesis on (P) that $\sum ix_i\ge r\binom{n}{r}$, which is exactly what we aimed to show. □

Again, it is not too difficult to work with the program (P) directly, but we'll instead translate it into the form needed to apply our Weak Duality Theorem II.

Claim 31.11. *To prove the result, it suffices to show that the linear program (P') with variables $\mathbf{x}\in\mathbb{R}^n$ defined by*

$$\begin{aligned}\text{maximize } & \sum_i -ix_i\\ \text{subject to } & \sum_i -x_i\le-\binom{n}{r}\\ & \sum_i\binom{n}{i}^{-1}x_i\le 1\\ & \mathbf{x}\ge 0\end{aligned}$$

has optimum value at most $-r\binom{n}{r}$.

In the language of the Weak Duality Theorem Theorem 31.4, (P') has A the matrix whose first row is $(-1,-1,\dots,-1)$ and second row is $(\binom{n}{1}^{-1},\dots,\binom{n}{n}^{-1})$ with $\mathbf{b}=(-\binom{n}{r},1)$ and $\mathbf{c}=(-1,-2,\dots,-n)$. Weak Duality then immediately gives the following.

Claim 31.12. *To prove the result, it suffices to show that the linear program (D') with variables $y_1, y_2 \in \mathbb{R}$ defined by*

$$\begin{aligned} \text{minimize } & -\binom{n}{r} y_1 + y_2 \\ \text{subject to } & -y_1 + \binom{n}{i}^{-1} y_2 \geq -i \quad \forall i \\ & y_1, y_2 \geq 0 \end{aligned}$$

has optimum value at most $-r\binom{n}{r}$.

To prove this, we now only need to find some specific choice of y_1, y_2 satisfying the constraints of (D') which gives the desired value for the objective function. The short answer is that this will work out for $y_1 = \frac{r(n-2r+2)}{n-2r+1}$ and $y_2 = \frac{r\binom{n}{r}}{n-2r+1}$, as one can verify with some straightforward computations, proving the result. As per usual, we will opt for a slower approach in order to justify why we might consider looking at such a strange looking set of values.

To begin, it perhaps seems like amongst the constraints of the form $-y_1 + \binom{n}{i}^{-1} y_2 \geq -i$, the $i = r$ case might be the most relevant given the role r plays in the objective function. And indeed, this $i = r$ constraint is equivalent to saying $-\binom{n}{r} y_1 + y_2 \geq -r\binom{n}{r}$, and since our ultimate goal is to find y_1, y_2 such that $-\binom{n}{r} y_1 + y_2 \leq -r\binom{n}{r}$ (as this proves the optimum value is at most $-r\binom{n}{r}$), we see that to have any hope in succeeding we must choose y_1, y_2 such that $-\binom{n}{r} y_1 + y_2 = -r\binom{n}{r}$. This allows us to write y_1 as a function of y_2, and plugging this into each of the other constraints gives for all i the new constraint $(1 - \binom{n}{i}\binom{n}{r}^{-1}) y_2 \geq (r-i)\binom{n}{i}$, which after some rearranging is equivalent to having

$$\begin{cases} \binom{n}{r}^{-1} y_2 \geq \frac{(r-i)}{\binom{n}{r}\binom{n}{i}^{-1} - 1} & i < r, \\ \binom{n}{r}^{-1} y_2 \leq \frac{(r-i)}{\binom{n}{r}\binom{n}{i}^{-1} - 1} & i > r. \end{cases}$$

Heuristically the two extreme cases of these inequalities should be $i = r \pm 1$, and indeed one can show the following.

Claim 31.13. *If $n \geq 2r$, then for all $1 \leq i < r$ we have*

$$\frac{(r-i)}{\binom{n}{r}\binom{n}{i}^{-1} - 1} \leq \frac{1}{\binom{n}{r}\binom{n}{r-1}^{-1} - 1} = \frac{r}{n-2r+1},$$

and for all $r < i \leq n$ we have

$$\frac{(r-i)}{\binom{n}{r}\binom{n}{i}^{-1} - 1} \geq \frac{1}{\binom{n}{r}\binom{n}{r+1}^{-1} - 1} = \frac{n-r}{n-2r+1},$$

Proof. We prove only the $i < r$ case, the other proof being analogous. Observe that for any given i, the inequality $\frac{(r-i)}{\binom{n}{r}\binom{n}{i}^{-1} - 1} \leq \frac{1}{\binom{n}{r}\binom{n}{r-1}^{-1} - 1}$ is hardest to satisfy when $n \geq 2r$ is as small as possible, so it suffices to prove this when $n = 2r$. If $i = r - t$, then it is easy to check that

$$\binom{n}{r}\binom{n}{i}^{-1} - 1 = \frac{(n-r+1)(n-r+2)\cdots(n-r+t)}{r(r-1)\cdots(r-t+1)} - 1 \geq \frac{n-r+t}{r-t+1} - 1 = \frac{2t-1}{r-t+1}$$

with this inequality using $n-r+1+s \geq r-s$ for all $0 \leq s < t$ and the last equality using $n = 2r$. Since $r - i = t$, this implies

$$\frac{(r-i)}{\binom{n}{r}\binom{n}{i}^{-1} - 1} \leq \frac{t}{2t-1} \cdot (r-t+1) \leq 1 \cdot r = \frac{r}{n-2r+1},$$

completing the proof. $\square$

In total this claim implies that any choice of y_2 with $\frac{r\binom{n}{r}}{n-2r+1} \leq y_2 \leq \frac{(n-r)\binom{n}{r}}{n-2r+1}$ together with $y_1 = \binom{n}{r}^{-1} y_2 + r$ (both of which are easily checked to be non-negative) is a feasible solution to (D') that gives the value $-r\binom{n}{r}$ to the objective function, completing the proof of the theorem. $\square$

Before moving on, we note that there exist several other results from extremal set theory that can be proven using linear programming. For these problems, one can often take the variables x_i to correspond to the number of "objects" of "size i", where the exact definitions of these terms will depend on our problem. For example, in the proof above we took x_i to be the number of $A \in \mathcal{F}$ of size i. Another example due to Chowdhury [36] has the x_i denoting the number of pairs of elements of $[n]$ which have degree i in $\mathcal{F}$.

31.4 Strong Duality and Other Topics

Much more can be said about linear programming, and we again refer the reader to the book by Matoušek and Gärtner [115] for a more thorough treatment. Here we briefly discuss a few other topics of relevance to extremal combinatorics.

Strong Duality. Perhaps the most important concept about linear programming which has been omitted upto this point is the *Strong Duality Theorem*. Informally, this says that the optimal value of a dual program is not only an *upper bound* for the optimal value of the primal (which is the content of our Weak Duality Theorem), but is in fact *equal* to the optimal value of the primal.

Philosophically, this result says that every primal program has an "easy" proof of an optimal upper bound (namely, one can always find such a bound by taking a suitable linear combination of its constraints). On the practical side, strong duality says that for upper bounding primal programs, we only ever have to care about upper bounding its dual (that is, we never have to worry that our simple approach of using weak duality could give bounds that are far from optimal for our original program).

Linear Programming in Practice. Strong Duality shows that linear programs can be efficiently solved in theory. In fact, it turns out that linear programs can also be efficiently solved in *practice*, meaning that there exist fast algorithms for solving relatively large linear programs. Further, while general integer programs are difficult to solve in both theory and practice, there do exist reasonably fast algorithms that can be used to solve moderately sized integer programs.

With this in mind, if one has an extremal problem that can be phrased in terms of either a linear or integer program, then it might be possible to use these real-life efficient algorithms to find good constructions for some moderately large cases of the problem. In particular, Wagner [143] used exactly this approach to disprove around a dozen sporadic conjectures that were made in the field. Those interested in this direction of study might also enjoy Wagner's later work on using machine learning to come up with constructions in extremal combinatorics [144].

Other Types of Programs. After the statement of Theorem 31.4, we mentioned how one can introduce additional variables to the linear program in order to get it into the form of our Weak Duality Theorem, and in general the introduction of additional variables and constraints is a common trick used throughout linear programming. For example, say one has a problem which has linear constraint but one wants to minimize the objective function $|x_1 - x_2|$. This objective function doesn't fit the definition of a linear program, but it can be turned into a linear program by introducing a new variable y, adding the constraints $y \geq x_1 - x_2$ and $y \geq x_2 - x_1$, and then using y as the objective function we wish to minimize. Further examples of this kind can be found in [115].

While not every optimization problem can be transformed into a linear program, there are other "non-linear programs" which can also be used in extremal combinatorics. Perhaps the most common such programs are *semidefinite programs*, which very informally differs from linear programming in that they replace the vector of constraints $\mathbf{x}$ with a matrix of constraints X and replace the non-negativity condition $\mathbf{x} \geq 0$ with the hypothesis that X is positive semidefinite. The theoretical usage of semidefinite programs has found a number of applications in combinatorics (see e.g. [34]), but perhaps its most useful contribution to the field has been in its practical usage through the implementation of the method of *flag algebras* introduced by Razborov [127].

Very informally, the flag algebra method uses semidefinite programming to give systematic approaches for finding Cauchy-Schwarz style proofs to extremal combinatorics problems. This allows a computer to efficiently search for proofs which would normally be too long and complicated for any human to come up with on their own, and these proofs have been used to give the best known bounds for many problems in extremal combinatorics. Perhaps the most notable example of this is that of the Turán number of the 3-uniform clique $K_4^{(3)}$, which is famously conjectured to be asymptotically equal to $\frac{5}{9}\binom{n}{3} \approx .555\binom{n}{3}$. The best known upper bound of roughly $.5612\binom{n}{3}$ is due to Razborov [128] using the flag algebra method.

Part IX
Bibliography

References

[1] N. Alon. Combinatorial nullstellensatz. *Combinatorics, Probability and Computing*, 8(1-2):7–29, 1999.

[2] N. Alon, L. Babai, and H. Suzuki. Multilinear polynomials and frankl-ray-chaudhuri-wilson type intersection theorems. *Journal of Combinatorial Theory, Series A*, 58(2):165–180, 1991.

[3] N. Alon, M. Bucić, L. Sauermann, D. Zakharov, and O. Zamir. Essentially tight bounds for rainbow cycles in proper edge-colourings. *arXiv preprint arXiv:2309.04460*, 2023.

[4] N. Alon and S. Friedland. The maximum number of perfect matchings in graphs with a given degree sequence. *arXiv preprint arXiv:0803.2578*, 2008.

[5] N. Alon and Z. Füredi. Covering the cube by affine hyperplanes. *European journal of combinatorics*, 14(2):79–83, 1993.

[6] N. Alon, M. Krivelevich, and B. Sudakov. Turán numbers of bipartite graphs and related ramsey-type questions. *Combinatorics, Probability and Computing*, 12(5-6):477–494, 2003.

[7] N. Alon, M. Krivelevich, and B. Sudakov. Embedding nearly-spanning bounded degree trees. *Combinatorica*, 27(6):629–644, 2007.

[8] N. Alon and J. H. Spencer. *The probabilistic method*. John Wiley & Sons, 2016.

[9] R. Alweiss, S. Lovett, K. Wu, and J. Zhang. Improved bounds for the sunflower lemma. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 624–630, 2020.

[10] L. Babai and P. Frankl. *Linear algebra methods in combinatorics*. University of Chicago, 1988.

[11] S. Ball and O. Serra. Punctured combinatorial nullstellensätze. *Combinatorica*, 29(5):511–522, 2009.

[12] J. Balogh. Container Method notes.

[13] J. Balogh, B. Csaba, M. Pei, and W. Samotij. Large bounded degree trees in expanding graphs. *the electronic journal of combinatorics*, pages R6–R6, 2010.

[14] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *Journal of the American Mathematical Society*, 28(3):669–709, 2015.

[15] J. Balogh, R. Mycroft, and A. Treglown. A random version of sperner's theorem. *Journal of Combinatorial Theory, Series A*, 128:104–110, 2014.

[16] J. Balogh and Š. Petříčková. The number of the maximal triangle-free graphs. *Bulletin of the London Mathematical society*, 46(5):1003–1006, 2014.

[17] M. Bateman and N. Katz. New bounds on cap sets. *Journal of the American Mathematical Society*, 25(2):585–613, 2012.

[18] J. Beck. On size ramsey number of paths, trees, and circuits. i. *Journal of Graph Theory*, 7(1):115–129, 1983.

[19] T. Bell, S. Chueluecha, and L. Warnke. Note on sunflowers. *Discrete Mathematics*, 344(7):112367, 2021.

[20] P. Bennett, R. Cushman, and A. Dudek. Generalized ramsey numbers at the linear and quadratic thresholds. *arXiv preprint arXiv:2309.00182*, 2023.

[21] P. Bennett, M. Delcourt, L. Li, and L. Postle. On generalized ramsey numbers in the non-integral regime. *arXiv preprint arXiv:2212.10542*, 2022.

[22] P. Bennett, A. Dudek, and S. English. A random coloring process gives improved bounds for the erd\h {o} s-gy\'arf\'as problem on generalized ramsey numbers. *arXiv preprint arXiv:2212.06957*, 2022.

[23] A. Bishnoi, S. Boyadzhiyska, S. Das, and Y. d. Bakker. Covering grids with multiplicity. *arXiv preprint arXiv:2305.00825*, 2023.

[24] G. R. Blakley and P. Roy. A hölder type inequality for symmetric matrices with non-negative entries. *Proceedings of the American Mathematical Society*, 16(6):1244–1245, 1965.

[25] B. Bollobás. *Modern graph theory*, volume 184. Springer Science & Business Media, 2013.

[26] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *Journal of Combinatorial Theory, Series B*, 16(2):97–105, 1974.

[27] R. B. Boppana. A useful inequality for the binary entropy function. *arXiv preprint arXiv:2301.09664*, 2023.

[28] L. M. Brègman. Some properties of nonnegative matrices and their permanents. In *Doklady Akademii Nauk*, volume 211, pages 27–30. Russian Academy of Sciences, 1973.

[29] H. Bruhn and O. Schaudt. The journey of the union-closed sets conjecture. *Graphs and Combinatorics*, 31:2043–2074, 2015.

[30] B. Bukh. Random algebraic construction of extremal graphs. *arXiv preprint arXiv:1409.3856*, 2014.

[31] B. Bukh. Extremal graphs without exponentially-small bicliques. *arXiv preprint arXiv:2107.04167*, 2021.

[32] B. Bukh and D. Conlon. Rational exponents in extremal graph theory. *Journal of the European Mathematical Society*, 20(7):1747–1757, 2018.

[33] S. Cambie. Better bounds for the union-closed sets conjecture using the entropy approach. *arXiv preprint arXiv:2212.12500*, 2022.

[34] C. Carlson, A. Kolla, R. Li, N. Mani, B. Sudakov, and L. Trevisan. Lower bounds for max-cut in h-free graphs via semidefinite programming. *SIAM Journal on Discrete Mathematics*, 35(3):1557–1568, 2021.

[35] Z. Chase and S. Lovett. Approximate union closed conjecture. *arXiv preprint arXiv:2211.11689*, 2022.

[36] A. Chowdhury. On a conjecture of frankl and füredi. *Electronic Notes in Discrete Mathematics*, 38:259–263, 2011.

[37] F. R. Chung, Z. Füredi, R. L. Graham, and P. Seymour. On induced subgraphs of the cube. *Journal of Combinatorial Theory, Series A*, 49(1):180–187, 1988.

[38] A. Clifton and H. Huang. On almost k-covers of hypercubes. *Combinatorica*, 40(4):511–526, 2020.

[39] D. Conlon. Graphs with few paths of prescribed length between any two vertices. *Bulletin of the London Mathematical Society*, 51(6):1015–1021, 2019.

[40] D. Conlon and A. Ferber. Lower bounds for multicolor ramsey numbers. *Advances in Mathematics*, 378:107528, 2021.

[41] D. Conlon, J. Fox, and B. Sudakov. An approximate version of sidorenko's conjecture. *Geometric and Functional Analysis*, 20:1354–1366, 2010.

[42] D. Conlon, C. Pohoata, and D. Zakharov. Random multilinear maps and the erdos box problem. *arXiv preprint arXiv:2011.09024*, 2020.

[43] D. W. Cranston and D. B. West. An introduction to the discharging method via graph coloring. *Discrete Mathematics*, 340(4):766–793, 2017.

[44] E. Croot, V. F. Lev, and P. P. Pach. Progression-free sets in are exponentially small. *Annals of Mathematics*, pages 331–337, 2017.

[45] J. Cutler and A. Radcliffe. An entropy proof of the kahn-lovász theorem. *the electronic journal of combinatorics*, 18(1):P10, 2011.

[46] S. Das, C. Lee, and B. Sudakov. Rainbow turán problem for even cycles. *European Journal of Combinatorics*, 34(5):905–915, 2013.

[47] M. Delcourt and L. Postle. Finding an almost perfect matching in a hypergraph avoiding forbidden submatchings. *arXiv preprint arXiv:2204.08981*, 2022.

[48] D. Dellamonica Jr. The size-ramsey number of trees. *Random Structures & Algorithms*, 40(1):49–73, 2012.

[49] M. Deza, P. Erdös, and P. Frankl. Intersection properties of systems of finite sets. *Proceedings of the London Mathematical Society*, 3(2):369–384, 1978.

[50] M. Deza, P. Frankl, and N. M. Singhi. On functions of strength t. *Combinatorica*, 3(3-4):331–339, 1983.

[51] I. Dinur and E. Friedgut. Intersecting families are essentially contained in juntas. *Combinatorics, Probability and Computing*, 18(1-2):107–122, 2009.

[52] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

[53] Y. Edel. Extensions of generalized product caps. *Designs, Codes and Cryptography*, 31:5–14, 2004.

[54] J. S. Ellenberg and D. Gijswijt. On large subsets of with no three-term arithmetic progression. *Annals of Mathematics*, pages 339–343, 2017.

[55] P. Erdös. Some remarks on the theory of graphs. *Bulletin of the American Mathematical Society*, 53(4):292–294, 1947.

[56] P. Erdos. Problems and results in combinatorial analysis. In *Colloq. Internat. Theor. Combin. Rome*, pages 3–17, 1973.

[57] P. Erdős, R. J. Faudree, C. C. Rousseau, and R. H. Schelp. The size ramsey number. *Periodica Mathematica Hungarica*, 9(1-2):145–161, 1978.

[58] P. Erdős and A. Gyárfás. A variant of the classical ramsey problem. *Combinatorica*, 17:459–467, 1997.

[59] P. Erdos and H. Hanani. On a limit theorem in combinatorial analysis. *Publ. Math. Debrecen*, 10(10-13):2–2, 1963.

[60] P. Erdos, D. J. Kleitman, and B. L. Rothschild. Asymptotic enumeration of kn-free graphs. *Colloquio Internazionale sulle Teorie Combinatorie (Rome, 1973)*, 2(17):19–27, 1976.

[61] P. Erdös and E. Szemerédi. Combinatorial properties of systems of sets. *Journal of Combinatorial Theory, Series A*, 24(3):308–313, 1978.

[62] R. J. Faudree and M. Simonovits. On a class of degenerate extremal graph problems. *Combinatorica*, 3:83–93, 1983.

[63] J. Fox and B. Sudakov. Dependent random choice. *Random Structures & Algorithms*, 38(1-2):68–99, 2011.

[64] P. Frankl. Extremal set systems. In *Handbook of combinatorics (vol. 2)*, pages 1293–1329. 1996.

[65] P. Frankl and Z. Füredi. Forbidding just one intersection. *Journal of Combinatorial Theory, Series A*, 39(2):160–176, 1985.

[66] P. Frankl and N. Tokushige. *Extremal problems for finite sets*, volume 86. American Mathematical Soc., 2018.

[67] P. Frankl and R. M. Wilson. Intersection theorems with geometric consequences. *Combinatorica*, 1:357–368, 1981.

[68] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *arXiv preprint arXiv:1910.13433*, 2019.

[69] E. Friedgut. On the measure of intersecting families, uniqueness and stability. *Combinatorica*, 28(5):503–528, 2008.

[70] J. Friedman and N. Pippenger. Expanding graphs contain all small trees. *Combinatorica*, 7:71–76, 1987.

[71] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2016.

[72] Z. Füredi. On a turán type problem of erdös. *Combinatorica*, 11(1):75–79, 1991.

[73] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdős Centennial*, pages 169–264. Springer, 2013.

[74] D. Galvin. Three tutorial lectures on entropy and counting. *arXiv preprint arXiv:1406.7872*, 2014.

[75] D. Galvin. Counting colorings of a regular graph. *Graphs and Combinatorics*, 31(3):629–638, 2015.

[76] D. Galvin and P. Tetali. On weighted graph homomorphisms. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 63:97–104, 2004.

[77] J. Gilmer. A constant lower bound for the union-closed sets conjecture. *arXiv preprint arXiv:2211.09055*, 2022.

[78] R. Glebov. *On Hamilton cycles and other spanning structures*. PhD thesis, 2013.

[79] S. Glock, F. Joos, J. Kim, M. Kühn, and L. Lichev. Conflict-free hypergraph matchings. *Journal of the London Mathematical Society*, 109(5):e12899, 2024.

[80] S. Glock, D. Kühn, A. Lo, and D. Osthus. The existence of designs via iterative absorption: hypergraph f-designs for arbitrary f. *arXiv preprint arXiv:1611.06827*, 2016.

[81] S. Glock, D. Kühn, A. Lo, and D. Osthus. On a conjecture of erdős on locally sparse steiner triple systems. *Combinatorica*, 40(3):363–403, 2020.

[82] C. Gotsman and N. Linial. The equivalence of two problems on the cube. *Journal of Combinatorial Theory, Series A*, 61(1):142–146, 1992.

[83] W. Gowers. Topics in combinatorics. *Notes taken by Paul Russell)*, 2004.

[84] J. Grochow. New applications of the polynomial method: the cap set conjecture and beyond. *Bulletin of the American Mathematical Society*, 56(1):29–64, 2019.

[85] D. S. Gunderson, V. Rödl, and A. Sidorenko. Extremal problems for sets forming boolean algebras and complete partite hypergraphs. *Journal of Combinatorial Theory, Series A*, 88(2):342–367, 1999.

[86] P. E. Haxell. Tree embeddings. *Journal of Graph Theory*, 36(3):121–130, 2001.

[87] A. Heckel. Random triangles in random graphs. *Random Structures & Algorithms*, 2018.

[88] A. Heckel and O. Riordan. How does the chromatic number of a random graph vary? *arXiv preprint arXiv:2103.14014*, 2021.

[89] M. Hochberg. Restricted sperner families and s-systems. *Discrete Mathematics*, 3(4):359–364, 1972.

[90] H. Huang. Induced subgraphs of hypercubes and a proof of the sensitivity conjecture. *Annals of Mathematics*, 190(3):949–955, 2019.

[91] O. Janzer. Rainbow tur\'an number of even cycles, repeated patterns and blow-ups of cycles. *arXiv preprint arXiv:2006.01062*, 2020.

[92] O. Janzer and B. Sudakov. On the turán number of the hypercube. In *Forum of Mathematics, Sigma*, volume 12, page e38. Cambridge University Press, 2024.

[93] T. Jiang, S. Letzter, A. Methuku, and L. Yepremyan. Rainbow clique subdivisions and blow-ups. *arXiv preprint arXiv:2108.08814*, 2021.

[94] T. Jiang and R. Seiver. Turán numbers of subdivided graphs. *SIAM Journal on Discrete Mathematics*, 26(3):1238–1255, 2012.

[95] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[96] J. Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(3):219–237, 2001.

[97] J. Kahn, B. Narayanan, and J. Park. The threshold for the square of a hamilton cycle. *arXiv preprint arXiv:2010.08592*, 2020.

[98] P. Keevash. The existence of designs. *arXiv preprint arXiv:1401.3665*, 2014.

[99] P. Keevash, D. Mubayi, B. Sudakov, and J. Verstraëte. Rainbow turán problems. *Combinatorics, Probability and Computing*, 16(1):109–126, 2007.

[100] J. Kim, J. Lee, H. Liu, and T. Tran. Rainbow cycles in properly edge-colored graphs. *Combinatorica*, pages 1–11, 2024.

[101] J. H. Kim. The ramsey number r (3, t) has order of magnitude t2/log t. *Random Structures & Algorithms*, 7(3):173–207, 1995.

[102] D. Kleitman. A conjecture of erdős-katona on commensurable pairs among subsets of an n-set. In *Theory of Graphs, Proc. Colloq. held at Tihany, Hungary*, pages 187–207, 1966.

[103] D. Kleitman. On dedekind's problem: the number of monotone boolean functions. *Proceedings of the American Mathematical Society*, 21(3):677–682, 1969.

[104] D. J. Kleitman and E. Milner. On the average size of the sets in a sperner family. *Discrete Mathematics*, 6(2):141–147, 1973.

[105] E. Knill. Graph generated union-closed families of sets. *arXiv preprint math/9409215*, 229, 1994.

[106] A. Kupavskii and D. Zakharov. Spread approximations for forbidden intersections problems. *arXiv preprint arXiv:2203.13379*, 2022.

[107] W. Kuszmaul and Q. Qi. The multiplicative version of azuma's inequality, with an application to contention analysis. *arXiv preprint arXiv:2102.05077*, 2021.

[108] M. Kwan, A. Sah, M. Sawhney, and M. Simkin. High-girth steiner triple systems. *arXiv preprint arXiv:2201.04554*, 2022.

[109] L. H. Loomis and H. Whitney. An inequality related to the isoperimetric inequality. 1949.

[110] L. Lovász. *Combinatorial problems and exercises*, volume 361. American Mathematical Soc., 2007.

[111] W. Mader. Homomorphieeigenschaften und mittlere kantendichte von graphen. *Mathematische Annalen*, 174(4):265–268, 1967.

[112] T. C. Martinez. The slice rank polynomial method. 2021.

[113] J. Matoušek. *Thirty-three miniatures: Mathematical and Algorithmic applications of Linear Algebra*. American Mathematical Society Providence, RI, 2010.

[114] J. Matoušek. *Lectures on discrete geometry*, volume 212. Springer Science & Business Media, 2013.

[115] J. Matoušek and B. Gärtner. *Understanding and using linear programming*, volume 1. Springer, 2007.

[116] C. McDiarmid. General first-passage percolation. *Advances in Applied Probability*, 15(1):149–161, 1983.

[117] M. Michałek. A short proof of combinatorial nullstellensatz. *The American Mathematical Monthly*, 117(9):821–823, 2010.

[118] R. Montgomery. Spanning trees in random graphs. *Advances in Mathematics*, 356:106793, 2019.

[119] J. W. Moon and L. Moser. On cliques in graphs. *Israel journal of Mathematics*, 3:23–28, 1965.

[120] D. Mubayi and J. Verstraëte. A survey of turán problems for expansions. *Recent trends in combinatorics*, pages 117–143, 2016.

[121] D. Mubayi and Y. Zhao. Forbidding complete hypergraphs as traces. *Graphs and Combinatorics*, 23(6):667–679, 2007.

[122] E. Naslund and W. Sawin. Upper bounds for sunflower-free sets. In *Forum of Mathematics, Sigma*, volume 5, page e15. Cambridge University Press, 2017.

[123] J. Radhakrishnan. An entropy proof of bregman's theorem. *Journal of combinatorial theory, Series A*, 77(1):161–164, 1997.

[124] A. Rao. Coding for sunflowers. *arXiv preprint arXiv:1909.04774*, 2019.

[125] A. Rao. Sunflowers: from soil to oil. *Bulletin of the American Mathematical Society*, 60(1):29–38, 2023.

[126] D. K. Ray-Chaudhuri and R. M. Wilson. On t-designs. 1975.

[127] A. A. Razborov. Flag algebras. *The Journal of Symbolic Logic*, 72(4):1239–1282, 2007.

[128] A. A. Razborov. On 3-hypergraphs with forbidden 4-vertex configurations. *SIAM Journal on Discrete Mathematics*, 24(3):946–963, 2010.

[129] O. Riordan. Random cliques in random graphs. *arXiv preprint arXiv:1802.01948*, 2018.

[130] V. Rödl. On a packing and covering problem. *European Journal of Combinatorics*, 6(1):69–78, 1985.

[131] D. Romik. *The surprising mathematics of longest increasing subsequences*. Number 4. Cambridge University Press, 2015.

[132] A. Sah, M. Sawhney, D. Stoner, and Y. Zhao. The number of independent sets in an irregular graph. *Journal of Combinatorial Theory, Series B*, 138:172–195, 2019.

[133] W. Sawin. An improved lower bound for multicolor ramsey numbers and a problem of erdős. *Journal of Combinatorial Theory, Series A*, 188:105579, 2022.

[134] J. B. Shearer. On a problem of spencer. *Combinatorica*, 5(3):241–245, 1985.

[135] A. Sheffer. Polynomial methods and incidence theory.

[136] M. Simonovits and E. Szemerédi. Embedding graphs into larger graphs: results, methods, and problems. In *Building Bridges II*, pages 445–592. Springer, 2019.

[137] J. Spencer. Ramsey's theorem—a new lower bound. *Journal of Combinatorial Theory, Series A*, 18(1):108–115, 1975.

[138] J. Spencer. *Asymptopia*, volume 71. American Mathematical Soc., 2014.

[139] S. Spiro. A smoother notion of spread hypergraphs. *arXiv preprint arXiv:2106.11882*, 2021.

[140] T. Tao. A symmetric formulation of the croot-lev-pach-ellenberg-gijswijt capset bound. *blog post*, 2016.

[141] T. Tao and V. H. Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.

[142] I. Tomon. Robust (rainbow) subdivisions and simplicial cycles. *arXiv preprint arXiv:2201.12309*, 2022.

[143] A. Z. Wagner. Refuting conjectures in extremal combinatorics via linear programming. *Journal of Combinatorial Theory, Series A*, 169:105130, 2020.

[144] A. Z. Wagner. Constructions in combinatorics via neural networks. *arXiv preprint arXiv:2104.14516*, 2021.

[145] Y. Wigderson. An improved lower bound on multicolor ramsey numbers. *Proceedings of the American Mathematical Society*, 149(6):2371–2374, 2021.

[146] R. M. Wilson. The exact bound in the erdös-ko-rado theorem. *Combinatorica*, 4(2):247–257, 1984.

[147] Y. Zhao. The number of independent sets in a regular graph. *Combinatorics, Probability and Computing*, 19(2):315–320, 2010.

[148] Y. Zhao. The bipartite swapping trick on graph homomorphisms. *SIAM Journal on Discrete Mathematics*, 25(2):660–680, 2011.

[149] Y. Zhao. *Graph theory and additive combinatorics: exploring structure and randomness.* Cambridge University Press, 2023.