Math 492/529: Extremal Combinatorics

Jonathan A. Noel

University of Victoria
First Term, 2022/2023

December 30, 2022

Contents

Preface

> It is not enough to be in the right place at the right time. You should also have an open mind at the right time.
>
> *Paul Erdős*, lamenting the fact that he did not invent extremal graph theory (see [242]).

Figure 1: "My brain is open."

The topic of this course is extremal combinatorics, which is one of the most classical branches of combinatorial theory. Extremal combinatorics itself can be further divided into many distinct areas which can be traced back to the first half of the twentieth century, if not earlier. For instance,

- extremal set theory goes back at least as far as Sperner's Theorem from 1928 [247],
- extremal graph theory began with Mantel's Theorem from 1907 [193] and was first investigated in depth by Turán in 1941 [262],
- Ramsey Theory was popularised in a 1947 paper of Erdős [92] which was inspired by a result of Ramsey from 1929 [213] and
- extremal problems in arithemetic combinatorics grew from the work of van der Waerden in 1927 [265] and the Erdős–Turán Conjecture of 1936 [90].

While we will cover many of the "greatest hits" of extremal combinatorics from the past, one should not get the impression that this is a course on ancient history. On the contrary, extremal combinatorics is currently one of the most active branches of combinatorial theory, and many of the most exciting recent advances in extremal combinatorics harken back to the classical questions and theorems at its origin.

In designing this course, I have aimed for an emphasis on those topics which are most beautiful and are most relevant to modern research in the area. Fortunately, in extremal combinatorics, these two aims are almost completely aligned, and so very few compromises have had to be made. Of course, beauty is in the eye of the beholder, and evaluating "importance" is almost equally as subjective; therefore, it is inevitable that this course is biased towards my own interests. Some of the fundamental topics in extremal combinatorics are covered by other courses at UVic which some of you may have already taken (or may be taking now); in these cases, after building up the necessary background (i.e. definitions), we will usually take the topic in a different direction in order to offer a complementary perspective. We highlight some of these connections to other courses as they arise.

These notes have been divided into chapters. Each chapter brings together several topics of the same ilk. Each topic roughly fills one section. At the end of each chapter, there are two sections consisting of Exercises and Challenge Problems, which are described as follows:

- *Exercises* are problems which are chosen to check your understanding of the topics covered in the chapter. They mainly consist of questions which test your comprehension of the key definitions or your understanding of the main theorems and ability to apply them. Occasionally, these questions will require some new insights of limited depth. Some, but definitely not all, of these questions may be assigned for credit.

- *Challenge Problems* are questions which can be solved using the tools from this course (and perhaps a few that are not in the course) combined with a lot of ingenuity. These problems are for those who are keen to try something beyond the scope of the course or wish to "test their mettle." They may be assigned for bonus marks.

At the end of the notes, there is an appendix on asymptotic (Big O) notation. If you are not yet familiar with this notation, then it would be wise to familiarize yourself with it early, as it will be used throughout the course and you are likely to encounter it again in your mathematical studies. We will occasionally apply probabilistic arguments in this course. In all cases, these arguments will be of the most basic type, requiring nothing deeper than linearity of expectation and the "First Moment Method." I have also included an appendix on the basic facts from probability theory that we will use. There is also an appendix on some standard convexity inequalities which are applied in the notes. All three of these appendices are only included for the purposes of making these notes more-or-less self-contained.

There is also a fourth appendix consisting of "Further Remarks and References." This appendix has three main purposes:

- to provide additional context to the results discussed beyond the scope of the course,

- to post links to lecture videos and notes from similar courses at other universities around the world and

- to highlight the sources of some of the exercises and challenge problems from the notes.

I want to emphasize that it is **not necessary** for you to read anything in Appendix D in order to perform well in the course. I have only provided this information to supplement and enrich the course. For example, some students may find it useful to hear lectures or read notes given in a different style. Students who are interested in working on research projects in extremal combinatorics may benefit from seeing how the topics from these notes fit into the "bigger picture" of the subject. The sources of the exercises have been included mainly for my own future reference, but students may also like to use them to find additional exercises to practice (e.g. while preparing for the exam).

I hope that you enjoy the course as much as I enjoy teaching it.

Disclaimers and Acknowledgements

These notes are very much a work in progress and are likely to contain typos and mistakes. Please feel free to e-mail me with any corrections or questions. It is possible that these notes will change in the future to reflect the changing trends in extremal combinatorics.

There are several artistic figures in the notes that were made using Midjourney AI: `https://www.midjourney.com/home/`. I am grateful to Anita Liebenau for creating all of the nice-looking mathematical pictures. Any messy-looking mathematical pictures were drawn by me, by hand.

Thank you to all of the students who have taken this course in the past. The questions, comments and ideas that you have raised in class have inspired a few new exercises and examples and countless improvements to the exposition in these notes.

Chapter 1

Intersection, Complexity and Correlation

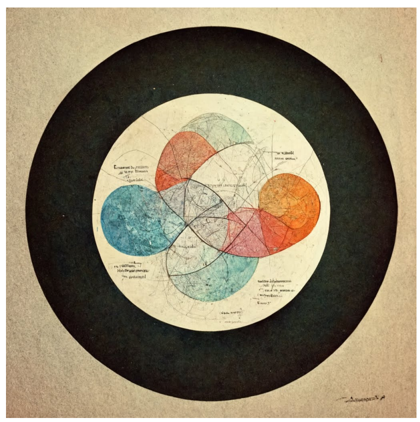


Figure 1.1: "Intersection Theorem."

The focus of this chapter, and the one which follows it, is on the combinatorial properties of collections of subsets of a finite set. Throughout, we let $[n]$ denote the set of the first n positive integers, i.e.

$$[n] := \{1, \ldots, n\}.$$

Given a set X, we let 2^X denote the *power set* of X; that is, the collection of all subsets of X. In particular,

$$2^{[n]} = \{A : A \subseteq [n]\}.$$

The power set of $[n]$ equipped with the partial order $\subseteq$ is sometimes referred to as the *boolean lattice*. Vaguely speaking, we will be considering questions of the following basic form:

> How large can a collection $\mathcal{F} \subseteq 2^{[n]}$ (which we call a *family* or *set system*) be, given that it satisfies certain constraints?

The focus of this chapter (with the exception of Section 1.4) is on extremal properties of set systems whose elements satisfy certain constraints on their intersections with one another.

1.1 Intersecting Families

A family $\mathcal{F}$ of sets is *intersecting* if $A \cap B \neq \emptyset$ for all $A, B \in \mathcal{F}$. It is not hard to determine the size of the largest intersecting family in $2^{[n]}$ (Hint: the answer is 2^{n-1}). The problem, however, becomes more interesting if we restrict to families of sets of the same cardinality. Consider the following definition.

Definition 1.1. Given a set X and integer k, define

$$\binom{X}{k} := \{A \subseteq X : |A| = k\}.$$

We refer to $\binom{X}{k}$ as the *kth level* of 2^X.

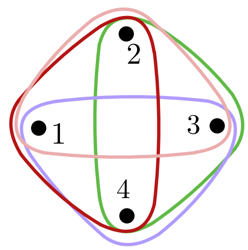


Figure 1.2: The four sets of $\binom{[4]}{3}$.

As you are probably aware,[1] the number of subsets of $[n]$ of cardinality k is equal to the kth binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Our focus in this section is on determining the maximum cardinality of an intersecting family $\mathcal{F} \subseteq \binom{[n]}{k}$. Let us analyze a few examples.

Example 1.2. If $k > n/2$, then any two sets in $\binom{[n]}{k}$ have non-empty intersection by the Pigeonhole Principle, and so the answer is simply $\binom{n}{k}$.

Example 1.3. Suppose that $k \leq n/2$ and define

$$\mathcal{F} := \{A \subseteq [n] : 1 \in A\}.$$

Clearly, $\mathcal{F}$ is intersecting. Thus, there exists an intersecting family of cardinality $\binom{n-1}{k-1}$.

Example 1.4. Suppose that $n = 2k$ and let $\mathcal{F} \subseteq \binom{n}{k}$ be any collection such that $\mathcal{F}$ contains exactly one of A or $[n] \setminus A$ for every $A \in \binom{[n]}{k}$ (note that both of these sets have size k, since $n = 2k$). Then it is not hard to see that $\mathcal{F}$ is intersecting and satisfies

$$|\mathcal{F}| = \frac{1}{2}\binom{n}{k} = \frac{1}{2}\binom{2k}{k} = \binom{2k-1}{k-1} = \binom{n-1}{k-1}.$$

[1] Binomial coefficients seem to usually be covered in the courses Math 151, 222 and 422 and frequently used in Math 322 and 423 at UVic. See, e.g., [68, Section 1.2].

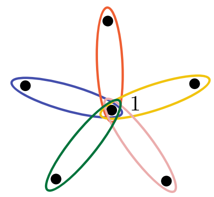


Figure 1.3: The construction in Example 1.3 in the case $n = 6$ and $k = 2$.

(the equality $\frac{1}{2}\binom{2k}{k} = \binom{2k-1}{k-1}$ is by Pascal's Formula[2] and the fact that $\binom{a}{b} = \binom{a}{b-a}$). Thus, this is an alternative construction giving the same bound as Example 1.3 (but only in the special case $n = 2k$).

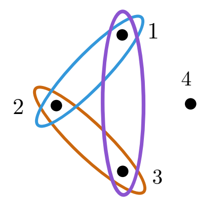


Figure 1.4: One possible construction from Example 1.4 in the case $n = 4$ and $k = 2$.

As it turns out, the constructions in Examples 1.3 and 1.4 are optimal.

Theorem 1.5 (Erdős–Ko–Rado Theorem [83])**.** *If n and k are integers such that $n \geq 2k$, then every intersecting family $\mathcal{F} \subseteq \binom{[n]}{k}$ satisfies $|\mathcal{F}| \leq \binom{n-1}{k-1}$.*

Proof. Let $\mathbb{Z}_n$ denote the integers modulo n; that is, $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z}$. An *interval* in $\mathbb{Z}_n$ is a set of the form $\{i, i+1, \ldots, i+t\}$ for $0 \leq i, t \leq n-1$, where addition is viewed modulo n.

Let $f : [n] \to \mathbb{Z}_n$ be a random bijection. For any set $A \subseteq [n]$ and an interval I in $\mathbb{Z}_n$, each of cardinality exactly k, the probability that $f(A) = I$ is precisely

$$\frac{k!(n-k)!}{n!} = \frac{1}{\binom{n}{k}}.$$

There are exactly n intervals of length k in $\mathbb{Z}_n$. Thus, by linearity of expectation,[3] the expected number of sets $A \in \mathcal{F}$ mapped to an interval under f is

$$\sum_{A \in \mathcal{F}} \left(\sum_{I \text{ an interval in } \mathbb{Z}_n} \mathbb{P}(f(A) = I) \right) = \frac{|\mathcal{F}| \cdot n}{\binom{n}{k}}.$$

Next, let us prove that no bijection $f : [n] \to \mathbb{Z}_n$ can map more than k sets of $\mathcal{F}$ to intervals. Here is the reason. Suppose that $A \in \mathcal{F}$ such that $f(A)$ is the interval $\{i, i+1, \ldots, i+k-1\}$. If B

[2]See the notes of Math 422 at UVic [68, p. 6].
[3]See Appendix B.

is another set of $\mathcal{F}$, then, since $A \cap B \neq \emptyset$, we have $f(A) \cap f(B) \neq \emptyset$. If $f(B)$ is an interval, then $f(B)$ must start at one of the points $i+1, \ldots, i+k-1$ or end at one of the points $i, \ldots, i+k-2$; note that it cannot start at i or end at $i+k-1$ because $A \neq B$. Moreover, since $n \geq 2k$, the interval $f(B)$ cannot both start and end in $f(A)$. Therefore, each set $B \in \mathcal{F}$ with $B \neq A$ which is mapped to an interval "cuts" the set A into two intervals (the elements of $B \cap A$ and $A \setminus B$). No two different sets $B \in \mathcal{F}$ can cut A in the same way, as this would imply that those two sets are either the same or are disjoint (again, we are using $n \geq 2k$). Thus, there are at most $k-1$ sets in $\mathcal{F}$ apart from A which map to an interval; so, there is a total of at most k such sets overall.

Putting this all together,

$$\frac{|\mathcal{F}| \cdot n}{\binom{n}{k}} \leq k$$

which implies that

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

and we are done. □

Later, in Section 2.3, we will see a second proof of the Erdős–Ko–Rado Theorem using the powerful Kruskal–Katona Theorem.

1.2 The Sunflower Lemma

Figure 1.5: "Sunflower."

In the previous section, we proved that, if $\mathcal{F} \subseteq \binom{[n]}{k}$ is sufficiently large, then it is possible to find two sets in $\mathcal{F}$ with empty intersection. In this section, our aim will be to find a large subcollection of $\mathcal{F}$ such that any pair of sets in this subcollection have *the same* intersection. This leads us to the following definition.

Definition 1.6. For $p \geq 2$, a *sunflower* with p *petals* is a collection $\{F_1, \ldots, F_p\}$ of p sets such that, for any $1 \leq i < j \leq p$ and $1 \leq i' < j' \leq p$,

$$F_i \cap F_j = F_{i'} \cap F_{j'}.$$

That is, all pairs of distinct sets in the collection have a common intersection. The *core* of the sunflower is defined to be the common intersection; i.e. the set $F_1 \cap F_2$.

In the literature, sunflowers are sometimes called Δ-*systems*. The following result of Erdős and Rado [84] implies that any sufficiently large collection of sets of bounded cardinality contains a sunflower.

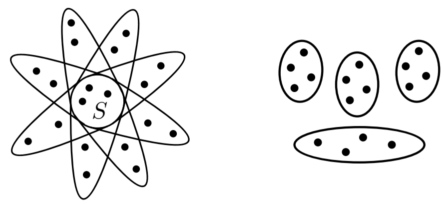


Figure 1.6: Two sunflowers. The one on the left has a core S of cardinality three and eight petals. The (sad looking) one on the right has an empty core and four petals.

Theorem 1.7 (Sunflower Lemma [84]). *For $k \geq 1$, if $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ such that $|\mathcal{F}| > k!(p-1)^k$, then $\mathcal{F}$ contains a sunflower with p petals.*

Proof. We proceed by induction on k. In the case $k = 1$, we have that $\mathcal{F}$ is a collection of at least p distinct single-element sets, and so any p of them form a sunflower with core $\emptyset$.

Now, suppose that $k \geq 2$. Let $\mathcal{A} = \{A_1, \ldots, A_t\}$ be a maximal collection of pairwise disjoint elements of $\mathcal{F}$. If $t \geq p$, then we are done, as the sets in $\mathcal{A}$ form a sunflower with core $\emptyset$. So, we assume that $t \leq p - 1$. Define

$$X := \bigcup_{i=1}^{t} A_i.$$

Since every set in $\mathcal{F}$ has cardinality k and $t \leq p - 1$, we have that $|X| \leq (p-1)k$. By maximality of $\mathcal{A}$, every element of $\mathcal{F}$ intersects X. Therefore, by the Pigeonhole Principle,[4] there is a point $x \in X$ which is contained in at least

$$\frac{|\mathcal{F}|}{|X|} > \frac{k!(p-1)^k}{(p-1)k} = (k-1)!(p-1)^{k-1}$$

sets of $\mathcal{F}$. Consider the system

$$\mathcal{F}' := \{F \setminus \{x\} : F \in \mathcal{F} \text{ and } x \in F\}.$$

Then $\mathcal{F}'$ is a collection of more than $(k-1)!(p-1)^{k-1}$ sets of cardinality $k-1$; therefore, it has a sunflower with p petals by the induction hypothesis. Adding the element x back into each set of this sunflower gives us a sunflower with p petals in $\mathcal{F}$. $\square$

One interesting thing to mention is that, unlike the Erdős–Ko–Rado Theorem in the previous section (and many other results covered in these notes), the Sunflower Lemma is *not* tight. That is, for most values of k and p, there does not exist a collection $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ with $|\mathcal{F}| = k!(p-1)^k$ such that $\mathcal{F}$ does not contain a sunflower with p petals. In Exercise 1.19, we describe a lower bound construction which is rather far from matching the upper bound in Theorem 1.7.

[4]See the course notes for Math 422 at UVic [68, Chapter 5].

1.3 VC-Dimension

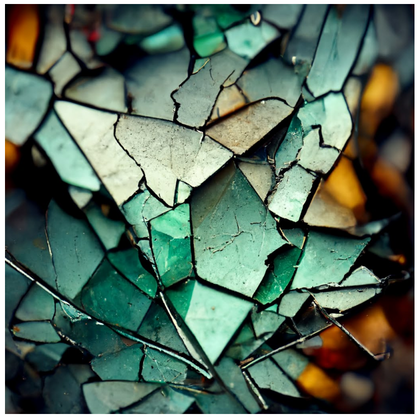


Figure 1.7: “Shattered.”

How does one measure the “complexity” of a set system? Roughly speaking, one may say that a set system is very complex is to say that it contains a “copy” of $2^{[k]}$ within it for some reasonably large value of k. One way of making this more rigorous is with the notion of VC-dimension.

Definition 1.8. Given a set X and a collection $\mathcal{F} \subseteq 2^X$, we say that a set $S \subseteq X$ is *shattered* by $\mathcal{F}$ if

$$\{F \cap S : F \in \mathcal{F}\} = 2^S.$$

In other words, a set S is shattered if every subset of S can be “represented” as the intersection of S with a set in $\mathcal{F}$.

Definition 1.9. The *VC-dimension* of a family $\mathcal{F} \subseteq 2^X$ is the maximum cardinality of a set $S \subseteq X$ which is shattered by $\mathcal{F}$; if arbitrarily large sets are shattered by $\mathcal{F}$, then the VC-dimension is ∞.

Example 1.10. Suppose that $\mathcal{F}$ is the collection of all closed intervals $[a, b]$ contained in $\mathbb{R}$. We claim that the VC-dimension of $\mathcal{F}$ is 2. Consider $S = \{1, 2\}$. We have that

$$S \cap [3, 100] = \emptyset, \qquad S \cap [1.5, 5] = \{2\},$$

$$S \cap [0, 1.5] = \{1\}, \qquad S \cap [0, 3] = \{1, 2\}$$

and so S is shattered by $\mathcal{F}$; this certifies that the VC-dimension is at least two. Now, given three reals $x < y < z$, there is no interval $[a, b]$ which contains x and z but not y. Thus, no set of cardinality three (or greater) can be shattered.

Example 1.11. Suppose that $\mathcal{F}$ is the collection of all rectangles of the form $[a, b] \times [c, d]$ in $\mathbb{R}^2$; i.e. *axis-aligned* rectangles. We claim that the VC-dimension is 4. Take the set $S = \{(-1, 0), (1, 0), (0, 1), (0, -1)\}$. If you play around for a bit, you can see that, for any subset of S, there is an axis-aligned rectangle which intersects S precisely on that subset; see Figure 1.3.

Now, take any set S of five points in $\mathbb{R}^2$. We can let x, y, z and w be the leftmost, rightmost, bottommost and topmost elements of this set (some of them may be equal). Now, it is not hard to see that there cannot be an axis-aligned rectangle which intersects S precisely on $\{x, y, z, w\}$.

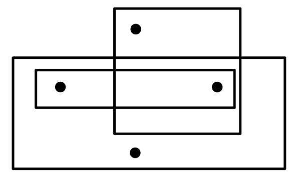


Figure 1.8: Some examples of rectangles used to shatter the set $\{(-1,0),(1,0),(0,1),(0,-1)\}$ in Example 1.11.

A natural question is, how large must a collection $\mathcal{F} \subseteq 2^{[n]}$ be in order to guarantee that it has VC-dimension at least k?

Example 1.12. Consider the family $\mathcal{F} = \bigcup_{i=0}^{k-1} \binom{[n]}{i}$. Clearly, $\mathcal{F}$ cannot shatter any set of cardinality k as $\mathcal{F}$ does not even contain a set of this size.

As it turns out, Example 1.12 is optimal, as the next result shows.

Theorem 1.13 (Sauer–Shelah Lemma [227, 237])**.** *If $\mathcal{F} \subseteq 2^{[n]}$ such that*

$$|\mathcal{F}| > \sum_{i=0}^{k-1} \binom{n}{i},$$

then the VC-dimension of $\mathcal{F}$ is at least k.

Proof. We will actually prove the stronger statement that every non-empty collection $\mathcal{F} \subseteq 2^{[n]}$ shatters at least $|\mathcal{F}|$ distinct subsets of $[n]$. Given this, the result will follow as there are at most $\sum_{i=0}^{k-1} \binom{n}{i}$ sets of cardinality less than k.

We proceed by induction on $|\mathcal{F}|$. Every non-empty family shatters the empty set, which proves the case $|\mathcal{F}| = 1$. Now, suppose that $|\mathcal{F}| \geq 2$ and let $\mathcal{S}$ be the collection of subsets of $[n]$ shattered by $\mathcal{F}$. Our goal is to prove that $|\mathcal{S}| \geq |\mathcal{F}|$.

Since $|\mathcal{F}| \geq 2$, we can let $F_1, F_2 \in \mathcal{F}$ such that $F_1 \setminus F_2 \neq \emptyset$. Choose $x \in F_1 \setminus F_2$ and let

$$\mathcal{F}_1 := \{F \in \mathcal{F} : x \in F\}$$

and

$$\mathcal{F}_2 := \{F \in \mathcal{F} : x \notin F\}.$$

For $i \in \{1,2\}$, let $\mathcal{S}_i$ be the collection of subsets of $[n]$ shattered by $\mathcal{F}_i$. The families $\mathcal{F}_1$ and $\mathcal{F}_2$ partition $\mathcal{F}$ and, by our choice of x, neither of them is empty. So, we can apply the inductive hypothesis to each of them. That is, $|\mathcal{S}_1| \geq |\mathcal{F}_1|$ and $|\mathcal{S}_2| \geq |\mathcal{F}_2|$.

Now, suppose that $S \subseteq [n]$ is shattered by either $\mathcal{F}_1$ or by $\mathcal{F}_2$. Then, by definition of shattering, we immediately get that S is also shattered by $\mathcal{F}$. Moreover, we observe that x cannot be an element of S; this is because every set in $\mathcal{F}_1$ contains x and no set in $\mathcal{F}_2$ contains x. Therefore,

$$\mathcal{S}_1 \cup \mathcal{S}_2 \subseteq \{S \in \mathcal{S} : x \notin S\}. \tag{1.14}$$

Of course, this is not quite enough to finish the proof, since $\mathcal{S}_1$ and $\mathcal{S}_2$ may have common elements. However, given any $S \in \mathcal{S}_1 \cap \mathcal{S}_2$, recall that $x \notin S$. Since both $\mathcal{F}_1$ and $\mathcal{F}_2$ shatter S, we get that $\mathcal{F}$ shatters the set $S \cup \{x\}$. Therefore,

$$\{S \cup \{x\} : S \in \mathcal{S}_1 \cap \mathcal{S}_2\} \subseteq \{S \in \mathcal{S} : x \in S\}. \tag{1.15}$$

Putting (1.14) and (1.15) together, and applying the inductive hypothesis, we get that

$$\begin{aligned} |\mathcal{S}| &= |\{S \in \mathcal{S} : x \notin S\}| + |\{S \in \mathcal{S} : x \in S\}| \\ &\geq |\mathcal{S}_1 \cup \mathcal{S}_2| + |\mathcal{S}_1 \cap \mathcal{S}_2| = |\mathcal{S}_1| + |\mathcal{S}_2| \\ &\geq |\mathcal{F}_1| + |\mathcal{F}_2| = |\mathcal{F}| \end{aligned}$$

and we are done. □

1.4 The Harris–Kleitman Inequality

Figure 1.9: “Correlation.”

Our goal in this section is to prove a beautiful correlation inequality, which is often referred to as Harris' Inequality or Kleitman's Inequality (depending on who you ask), and is a special case of several more general results that you may encounter in your mathematical career, such as the FKG Inequality and the Four Functions Theorem. On the surface, this inequality has little connection to the“intersection theorems” in the rest of this chapter. However, its relationship to the Erdős–Ko–Rado Theorem will come up several times in the exercises at the end of this chapter.

Say that a collection $\mathcal{F} \subseteq 2^{[n]}$ is a *downset* (or is *downward closed*) if whenever $A \in \mathcal{F}$ and $B \subseteq A$, it also holds that $B \in \mathcal{F}$. Analogously, $\mathcal{F}$ is an *upset* (or is *upward closed*) if whenever $A \in \mathcal{F}$ and $B \supseteq A$, it also holds that $B \in \mathcal{F}$. Given a pair $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$, how many elements must $\mathcal{A}$ and $\mathcal{B}$ have in common? Since both $\mathcal{A}$ and $\mathcal{B}$ are “biased” towards small subsets of $[n]$, one may expect them to overlap on a reasonably large number of elements; the following result confirms this intuition.

Theorem 1.16 (The Harris–Kleitman Inequality [130, 154]). *If $\mathcal{A}$ and $\mathcal{B}$ are both downsets in $2^{[n]}$, then*

$$\frac{|\mathcal{A} \cap \mathcal{B}|}{2^n} \geq \frac{|\mathcal{A}|}{2^n} \cdot \frac{|\mathcal{B}|}{2^n}.$$

Proof. The inequality that we are trying to prove can be equivalently written as

$$2^n |\mathcal{A} \cap \mathcal{B}| \leq |\mathcal{A}| \cdot |\mathcal{B}|.$$

We prove this by induction on n. The case $n = 1$ is easy, so let's consider $n \geq 2$. Define

$$\mathcal{A}^+ := \{A \subseteq [n-1] : A \cup \{n\} \in \mathcal{A}\}$$

and

$$\mathcal{A}^- := \{A \subseteq [n-1] : A \in \mathcal{A}\}.$$

Define $\mathcal{B}^+$ and $\mathcal{B}^-$ analogously. Note that $\mathcal{A}^+$ and $\mathcal{A}^-$ are both downsets in $2^{[n-1]}$. Moreover, since $\mathcal{A}$ is a downset, we have that $A \in \mathcal{A}$ whenever $A \cup \{n\} \in \mathcal{A}$; this implies that $\mathcal{A}^+ \subseteq \mathcal{A}^-$. Analogous properties hold for $\mathcal{B}^+$ and $\mathcal{B}^-$. Note that

$$2^n |\mathcal{A} \cap \mathcal{B}| = 2^n |\mathcal{A}^+ \cap \mathcal{B}^+| + 2^n |\mathcal{A}^- \cap \mathcal{B}^-|$$

By induction, this is at least

$$2|\mathcal{A}^+| \cdot |\mathcal{B}^+| + 2|\mathcal{A}^-| \cdot |\mathcal{B}^-|$$
$$= \left(|\mathcal{A}^+| + |\mathcal{A}^-|\right)\left(|\mathcal{B}^+| + |\mathcal{B}^-|\right) + \left(|\mathcal{A}^+| - |\mathcal{A}^-|\right)\left(|\mathcal{B}^+| - |\mathcal{B}^-|\right)$$

which, since $\mathcal{A}^+ \subseteq \mathcal{A}^-$ and $\mathcal{B}^+ \subseteq \mathcal{B}^-$, is at least

$$|\mathcal{A}| \cdot |\mathcal{B}|.$$

This completes the proof. □

Of course, the same is true for upsets.

Corollary 1.17. *If $\mathcal{A}$ and $\mathcal{B}$ are both upsets in $2^{[n]}$, then*

$$\frac{|\mathcal{A} \cap \mathcal{B}|}{2^n} \geq \frac{|\mathcal{A}|}{2^n} \cdot \frac{|\mathcal{B}|}{2^n}.$$

Proof. If $\mathcal{A}$ is an upset, then $\overline{\mathcal{A}} := \{[n] \setminus S : S \in \mathcal{A}\}$ is a downset. Also, $|\overline{\mathcal{A}}| = |\mathcal{A}|$ and $|\overline{\mathcal{A}} \cap \overline{\mathcal{B}}| = |\mathcal{A} \cap \mathcal{B}|$. The result now follows by Theorem 1.16. □

Next, let us show that the inequality in Theorem 1.16 is tight.

Example 1.18. Perhaps the simplest tight example for the Harris–Kleitman Inequality is to take $\mathcal{A} = \mathcal{B} = 2^{[n]}$; in this case, both sides of the inequality are equal to one.

For a more general family of examples, take S and T to be disjoint subsets of $[n]$ and let $\mathcal{A} := 2^{[n] \setminus S}$ and $\mathcal{B} := 2^{[n] \setminus T}$. Then $\mathcal{A} \cap \mathcal{B} = 2^{[n] \setminus (S \cup T)}$ and so, since S and T are disjoint, it has cardinality $2^{n - |S| - |T|}$. From this, we see that both sides of the inequality in Theorem 1.16 evaluate to $2^{-|S|-|T|}$ and so it is tight.

Perhaps the best way to think of the Harris–Kleitman Inequality is as a "correlation" inequality for downsets. That is, if one selects a element of $2^{[n]}$ uniformly at random, then the probability that it is contained in $\mathcal{A}$ is $|\mathcal{A}|/2^n$ and the probability that it is contained in $\mathcal{B}$ is $|\mathcal{B}|/2^n$. If these two events were independent, then the probability that it is contained in both $\mathcal{A}$ and $\mathcal{B}$, i.e. the probability that it is contained in $\mathcal{A} \cap \mathcal{B}$, would be the product of these two probabilities: $(|\mathcal{A}|/2^n) \cdot (|\mathcal{B}|/2^n)$. Of course, the actual probability of this event is given by $|\mathcal{A} \cap \mathcal{B}|/2^n$. Therefore, one can interpret this theorem as saying that, in the case that both collections are downsets, the events of a random set landing in $\mathcal{A}$ and landing in $\mathcal{B}$ are "positively correlated."

This allows us to derive inequalities relating the probability of the intersection of two events to the product of their individual probabilities. This is particularly useful when the probability of this intersection is difficult to compute exactly. For example, suppose that $\mathcal{F}$ is a random subset of $2^{[n]}$ obtained by including each $S \subseteq [n]$ with probability $1/2$, independently of all other such sets. Let E_1 be the event that $\mathcal{F}$ is intersecting and E_2 be the probability that $\mathcal{F}$ has VC-dimension at most $37 \cdot n^{1/2077}$. If the event E_1 is satisfied by $\mathcal{F}$, then it is also satisfied by every subcollection of $\mathcal{F}$; the same is true for E_2. Thus, by the Harris–Kleitman Inequality,

$$\mathbb{P}(E_1 \cap E_2) \geq \mathbb{P}(E_1)\mathbb{P}(E_2).$$

Of course, the probabilities of E_1 and E_2 may still be difficult to compute in this case, but they are probably easier to compute than the probability of $E_1 \cap E_2$.

1.5 Exercises

1.1 Prove that the maximum cardinality of an intersecting family $\mathcal{F} \subseteq 2^{[n]}$ is 2^{n-1}.

1.2 Prove that every intersecting family $\mathcal{F} \subseteq 2^{[n]}$ is contained in an intersecting family in $2^{[n]}$ of cardinality 2^{n-1}.

1.3 For $n \geq 5$, describe three non-isomorphic constructions of intersecting families $\mathcal{F} \subseteq 2^{[n]}$ of cardinality 2^{n-1}.

1.4 Prove that, if $\mathcal{F}_1, \ldots, \mathcal{F}_k$ are intersecting families in $2^{[n]}$, then

$$\left| \bigcup_{i=1}^{k} \mathcal{F}_i \right| \leq 2^n - 2^{n-k}.$$

Hint: The Harris–Kleitman Inequality and Exercise 1.1 may be useful.

1.5 For $n \leq 2k$, determine the size of the largest family $\mathcal{F} \subseteq \binom{[n]}{k}$ such that $A \cup B \neq [n]$ for all $A, B \in \mathcal{F}$.

1.6 Exactly how many intersecting families in $\binom{[n]}{2}$ are there?

1.7 (a) Given $x \in \{0,1,2\}^n$ and $1 \leq i \leq n$, let x_i denote the ith coordinate of x. Say that a set $F \subseteq \{0,1,2\}^n$ is *intersecting* if for any $x, y \in F$ there exists $1 \leq i \leq n$ such that $x_i = y_i \neq 0$. Determine the maximum size of a set $F \subseteq \{0,1,2\}^n$ that is intersecting.

(b) Say that a set $F \subseteq \{0,1,2\}^n$ is *weakly intersecting* if for any $x, y \in F$ there exists $1 \leq i \leq n$ such that $x_i \neq 0$ and $y_i \neq 0$. Suppose that n is odd and consider the set

$$B := \{x \in \{0,1,2\}^n : x_i \neq 0 \text{ for more than } n/2 \text{ indices } i\}.$$

Show that B is weakly intersecting, and that every weakly intersecting set has cardinality at most $|B|$.

Hint: Given a weakly intersecting $F \subseteq \{0,1,2\}^n$, consider the family $\mathcal{F} := \{\{i : x_i \neq 0\} : x \in F\} \subseteq 2^{[n]}$. It may be helpful to solve Exercises 1.1 and 1.2 first.

(c) Generalize the results of parts (a) and (b) to subsets of $\{0, 1, \ldots, r-1\}^n$ for all $r \geq 2$.

(d) Generalize the result in part (b) to all n (not just odd n).

1.8 An intersecting family $\mathcal{F} \subseteq \binom{[n]}{k}$ is *maximal* if $\mathcal{F} \cup \{S\}$ is not intersecting for any $S \in \binom{[n]}{k} \setminus \mathcal{F}$. For integers k and n, let $f_k(n)$ be the minimum size of a maximal intersecting family in $\binom{[n]}{k}$. For fixed k, does $f_k(n)$ tend to infinity as n tends to infinity?

1.9 For $n \geq k \geq 1$, suppose that $\mathcal{F} \subseteq \binom{[n]}{k}$ such that $|\mathcal{F}| \geq k+1$ and any $k+1$ sets in $\mathcal{F}$ have a common element. Prove that there exists $x \in [n]$ which is in every set in $\mathcal{F}$.

1.10 Find a collection $\mathcal{F} \subseteq 2^{[n]}$ of cardinality $2^{n-1}+1$ such that, for every $\mathcal{A} \subseteq \mathcal{F}$ such that $\bigcap_{S \in \mathcal{A}} S = \emptyset$, it holds that $\bigcup_{S \in \mathcal{A}} S = [n]$.

1.11 As in Exercise 1.10, let $\mathcal{F} \subseteq 2^{[n]}$ have the property that, for every $\mathcal{A} \subseteq \mathcal{F}$ such that $\bigcap_{S \in \mathcal{A}} S = \emptyset$, it holds that $\bigcup_{S \in \mathcal{A}} S = [n]$. Our goal in this exercise is to prove that $|\mathcal{F}| \leq 2^{n-1}+1$.

(a) Let G be a graph with vertex set $[n]$ defined by letting ij be an edge of G if every set in $\mathcal{F}$ intersects $\{i, j\}$. Prove that every vertex of G has degree at least one.

(b) Show that if G is a graph on vertex set $[n]$ of minimum degree at least one, then the number of sets $S \subseteq [n]$ such that $S \cap \{i,j\} \neq \emptyset$ for every $ij \in E(G)$ is at most $2^{n-1}+1$. Using this, conclude that $|\mathcal{F}| \leq 2^{n-1}+1$.

1.12 Suppose that $\mathcal{F} \subseteq 2^{[n]}$ such that $|S \cap T| \neq 1$ for all $S, T \in \mathcal{F}$. Prove that it is possible to colour $[n]$ with 2 colours such that there does not exist a set in $\mathcal{F}$ whose elements are all the same colour.

1.13 (a) A *matching* in a family $\mathcal{F} \subseteq 2^{[n]}$ is a collection $\mathcal{M} \subseteq \mathcal{F}$ such that $A \cap B = \emptyset$ for all $A, B \in \mathcal{M}$. Show that there exists a family $\mathcal{F} \subseteq 2^{[n]}$ of cardinality $\frac{3}{4}2^n$ which does not contain a matching of cardinality 3 such that, for every $X \in 2^{[n]} \setminus \mathcal{F}$, the family $\mathcal{F} \cup \{X\}$ contains a matching of cardinality 3.

(b) Generalize the construction from part (a) to matchings of any cardinality k such that $1 \leq k \leq n$. For different values of k, the cardinality of $\mathcal{F}$ should be different.

1.14 Let $\mathcal{A} \subseteq 2^{\mathbb{N}}$ be an intersecting family. Does there necessarily exist a finite set F such that $\{A \cap F : A \in \mathcal{A}\}$ is intersecting?

1.15 For $k \geq 1$ and $n \geq 2k$, prove that $\binom{[n]}{k}$ can be partitioned into $n-2k+2$ intersecting families.

1.16 A *permutation* of $[n]$ is a sequence $\pi = (\pi_1, \ldots, \pi_n)$ such that, for each $j \in [n]$, there is a unique $1 \leq i \leq n$ such that $\pi_i = j$. For example, $(1, 3, 2)$ is a permutation of $[3]$.

Say that permutations π and σ of $[n]$ *intersect* if there exists i such that $\pi_i = \sigma_i$. Determine the maximum size of an intersecting family of permutations of $[n]$.

1.17 A set system $\mathcal{F} \subseteq 2^{[n]}$ has *Property B* if the elements of $[n]$ can be coloured with two colours in such a way that no set $A \in \mathcal{F}$ is monochromatic.

(a) For $k \geq 1$, prove that $\binom{[2k-1]}{k}$ does not have Property B.

(b) Prove that, if there does not exist $A, B \in \mathcal{F}$ with $|A \cap B| = 1$, then $\mathcal{F}$ has Property B.

1.18 Determine the maximum size of a collection $\mathcal{F} \subseteq 2^{[n]}$ such that there does not exist $a \in [n]$ such that there are two different sets in $\mathcal{F}$ that contain a and two different sets in $\mathcal{F}$ that do not contain a.

1.19 Let $V_1, \ldots, V_k$ be pairwise disjoint sets, each of cardinality $p - 1$, and let $V = V_1 \cup \cdots \cup V_k$. Define

$$\mathcal{F} := \{F \subseteq V : |F \cap V_i| = 1 \text{ for } 1 \leq i \leq k\}.$$

Prove that $\mathcal{F}$ has cardinality $(p-1)^k$ and does not contain any sunflower with p petals.

1.20 Adapt the argument in the proof of Theorem 1.7 to prove a version of the Sunflower Lemma in the case that $\mathcal{F}$ is a multiset of sets in $\binom{\mathbb{N}}{k}$ (that is, $\mathcal{F}$ may contain multiple copies of the same element of $\binom{\mathbb{N}}{k}$); note that the bound on $|\mathcal{F}|$ should be different.

1.21 Let $\mathcal{F}$ be the collection of all convex subsets of $\mathbb{R}^2$. Show that $\mathcal{F}$ does not have bounded VC-dimension.

1.22 Let X be a set and let $\mathcal{F} \subseteq 2^X$. The *dual system* $\mathcal{F}^*$ is a subset of $2^{\mathcal{F}}$ defined as follows. For each $x \in X$, the dual system contains the set

$$\{F \in \mathcal{F} : x \in F\}.$$

(We ignore duplicate sets). Prove that, for every positive integer d, there exists $f(d)$ such that if $\mathcal{F}$ has VC-dimension at most d, then the VC-dimension of $\mathcal{F}^*$ is at most $f(d)$.

Hint: It may help to play around with the cases $d = 1$ and $d = 2$ before trying to generalize.

1.23 Let $\mathcal{F}$ be the collection of all closed half-planes in $\mathbb{R}^2$. Prove that $\mathcal{F}$ has VC-dimension three.

1.24 Let $\mathcal{F}$ be the collection of all $(n-1)$-dimensional hyperplanes in $\mathbb{R}^n$. Show that $\mathcal{F}$ has VC-dimension n.

1.25 Show that, if $\mathcal{A}, \mathcal{B} \subseteq 2^{[n]}$ such that $\mathcal{A}$ is a downset and $\mathcal{B}$ is an upset, then

$$\frac{|\mathcal{A} \cap \mathcal{B}|}{2^n} \leq \frac{|\mathcal{A}|}{2^n} \cdot \frac{|\mathcal{B}|}{2^n}.$$

1.26 (a) Let $\mathcal{F}$ be an upset and $\mathcal{G}$ be a downset in $2^{[n]}$. Define $\mathcal{F}^c := 2^{[n]} \setminus \mathcal{F}$ and $\mathcal{G}^c := 2^{[n]} \setminus \mathcal{G}$. Prove that

$$|\mathcal{F} \cap \mathcal{G}^c| \cdot |\mathcal{G} \cap \mathcal{F}^c| \geq |\mathcal{F} \cap \mathcal{G}| \cdot |\mathcal{F}^c \cap \mathcal{G}^c|.$$

(b) Let $\mathcal{F}, \mathcal{G}, \mathcal{F}^c$ and $\mathcal{G}^c$ be as in part (a). Using the result of part (a), or otherwise, prove that $|\mathcal{F} \cap \mathcal{G}|^{1/2} + |\mathcal{F}^c \cap \mathcal{G}^c|^{1/2} \leq 2^{n/2}$.

Hint 1: Square both sides.

Hint 2: $|\mathcal{F}^c \cap \mathcal{G}^c| = 2^n - |\mathcal{F} \cap \mathcal{G}^c| - |\mathcal{G} \cap \mathcal{F}^c| - |\mathcal{F} \cap \mathcal{G}|$.

(c) Prove that, if $\mathcal{A}$ and $\mathcal{B}$ are families in $2^{[n]}$ such that no set in $\mathcal{A}$ contains a set in $\mathcal{B}$ and no set in $\mathcal{B}$ contains a set in $\mathcal{A}$, then $|\mathcal{A}|^{1/2} + |\mathcal{B}|^{1/2} \leq 2^{n/2}$.

1.27 For $r \geq 2$, let $\mathcal{P}_r^n$ be the partially ordered set on $\{0, 1, \ldots, r-1\}^n$ where $(x_1, \ldots, x_n) \preceq (y_1, \ldots, y_n)$ if and only if $x_i \leq y_i$ for all $1 \leq i \leq n$. A *downset* in $\mathcal{P}_r^n$ is a set $A \subseteq \mathcal{P}_r^n$ such that if $x \in A$ and $y \preceq x$, then $y \in A$. Prove that, if A and B are downsets in $\mathcal{P}_r^n$, then

$$\frac{|A \cap B|}{r^n} \geq \frac{|A|}{r^n} \cdot \frac{|B|}{r^n}.$$

1.28 Let G be a random graph on n vertices obtained by including each edge in G with probability $1/2$, independently of all other edges. Let E be the event that G contains $n/3$ vertex-disjoint triangles and F be the event that G contains a complete graph with at least $3\log(n)$ vertices. Prove that

$$\mathbb{P}(E \cap F) \geq \mathbb{P}(E)\mathbb{P}(F).$$

1.29 Prove that the size of the largest family $\mathcal{F} \subseteq 2^{[n]}$ such that $\mathcal{F}$ is intersecting and $A \cup B \neq [n]$ for all $A, B \in \mathcal{F}$ is 2^{n-2}.

Hint: Use Exercise 1.25.

1.30 Watch the film *N is a Number*: `https://youtu.be/djQkj4SaUYs`.

1.6 Challenge Problems

1.1* For $k \geq 1$, let $\mathcal{A} \subseteq \binom{\mathbb{N}}{k}$ be an intersecting family. Does there necessarily exist a finite set F such that $\{A \cap F : A \in \mathcal{A}\}$ is intersecting?

1.2* Given a set $A \subseteq \{1, \ldots, n\}$ and $p \in (0, 1/2)$, let the *p-measure* of A be

$$\mu_p(A) := p^{|A|}(1-p)^{n-|A|}.$$

The *p-measure* of a family $\mathcal{F} \subseteq 2^{[n]}$ is then defined by

$$\mu_p(\mathcal{F}) := \sum_{A \in \mathcal{F}} \mu_p(A).$$

Prove that if $\mathcal{F} \subseteq 2^{[n]}$ is intersecting, then $\mu_p(\mathcal{F}) \leq p$.

1.3* The mayor of Redundantville has decided that the system for controlling the traffic light at the corner of Main Street and First Avenue is not complicated enough. In order to remedy this, he replaces the current system (a single switch with three positions for the colours green, yellow and red) with a new system that uses n switches, each of which has three possible positions. The new system has the property that, whenever you change all of the switches to a different position (at the same time), the colour of the light always changes. Prove that there is a single switch whose position completely determines the colour of the light; therefore, the mayor's efforts to complexify the system were in vain.

Chapter 2

Chains, Antichains and Shadows

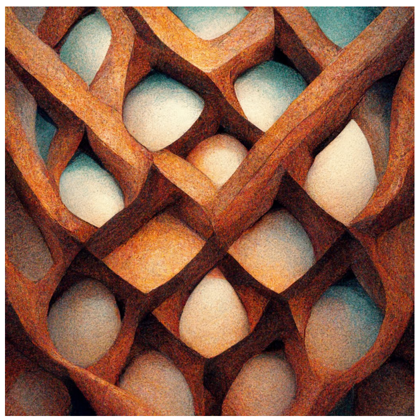


Figure 2.1: "Boolean lattice."

In this chapter, we continue our investigation of extremal properties of set systems. Here, the types of "constraints" that we focus on involve the subset relation on $2^{[n]}$.

2.1 Sperner's Theorem

We begin with a basic definition.

Definition 2.1. Say that a collection $\mathcal{A} \subseteq 2^{[n]}$ is an *antichain* if there does not exist $A, B \in \mathcal{A}$ such that $A \subsetneq B$.

Note that some people (especially matroid theorists) use the word *clutter* instead of antichain.

Example 2.2. Consider the case $n = 5$. One example of an antichain is

$$\mathcal{F} = \{\{1\}, \{2,3\}, \{2,4\}, \{3,4,5\}\}.$$

As you can check, there does not exist two distinct sets $A, B \in \mathcal{F}$ such that $A \subseteq B$.

Figure 2.2: The collection $\{\{1,2,4\},\{2,3,5,6\},\{5,6,7\}\}$ on the left is an antichain in $2^{[7]}$ because no set in the collection is contained in another set in the collection. The collection $\{\{1,2,3,4,5\},\{3,5,6,7\},\{3,6,7\}\}$ on the right is not an antichain in $2^{[7]}$ because $\{3,6,7\} \subseteq \{3,5,6,7\}$.

Our goal is to determine the cardinality of the largest antichain in $2^{[n]}$. Let us start with a simple example.

Example 2.3. If A and B are sets of the same cardinality, then it cannot hold that $A \subsetneq B$. Therefore, $\binom{[n]}{k}$ is an antichain in $2^{[n]}$ for every $0 \leq k \leq n$.

So, for every $0 \leq k \leq n$, there exists an antichain in $2^{[n]}$ of cardinality $\binom{n}{k}$. The largest possible antichain of this particular type is achieved when $k = \lfloor n/2 \rfloor$ or, equivalently, when $k = \lceil n/2 \rceil$.

Observation 2.4. *There exists an antichain $\mathcal{A} \subseteq 2^{[n]}$ such that $|\mathcal{A}| = \binom{n}{\lfloor n/2 \rfloor}$.*

The following theorem, known as *Sperner's Theorem*, asserts that the construction in Example 2.3 with $k = \lfloor n/2 \rfloor$ or $k = \lceil n/2 \rceil$ is, in fact, the largest possible in $2^{[n]}$. We state the theorem now and prove it later in this section.

Theorem 2.5 (Sperner's Theorem[1] [247])**.** *If $\mathcal{A} \subseteq 2^{[n]}$ is an antichain, then*

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

Remark 2.6. Antichains in $2^{[n]}$ are often called *Sperner families* or *Sperner systems* in honour of Sperner's Theorem [32, p. 10].

There are many different proofs of Sperner's Theorem. An approach using Hall's Theorem[2] on systems of distinct representatives (or, equivalently, matchings in bipartite graphs) can be found in the notes for Math 322 [192, Theorem 3.5]. Here, we provide an alternative proof which involves recursively constructing a so called "symmetric chain decomposition."

Definition 2.7. Say that a collection $\mathcal{C} \subseteq 2^{[n]}$ is a *chain* if either $A \subseteq B$ or $B \subseteq A$ for every pair $A, B \in \mathcal{C}$.

That is, a chain is the antithesis of an antichain.

[1] Sperner's Theorem is also covered in Math 322 at UVic [192, Theorem 3.5].
[2] Hall's Theorem is covered in Math 322 at UVic [192, Theorem 2.3].

Example 2.8. Let $n = 5$ and consider the collection

$$\mathcal{F} = \{\emptyset, \{1\}, \{1,4\}, \{1,2,4,5\}\}.$$

It is easy to check that $\mathcal{F}$ is a chain in $2^{[5]}$.

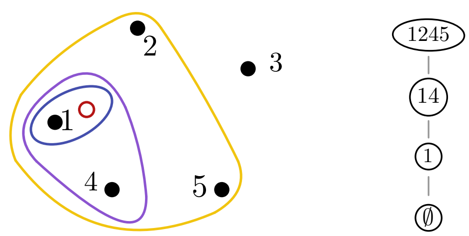


Figure 2.3: The chain in Example 2.8 (the empty set is depicted in red in the picture on the left).

As the names would suggest, a chain is the "antithesis" of an antichain.[3]

Observation 2.9. *If $\mathcal{A}, \mathcal{C} \subseteq 2^{[n]}$ such that $\mathcal{A}$ is an antichain and $\mathcal{C}$ is a chain, then $|\mathcal{A} \cap \mathcal{C}| \leq 1$.*

Proof. Suppose that A and B are distinct elements of $\mathcal{A} \cap \mathcal{C}$. Since they are both in $\mathcal{C}$, we have either $A \subseteq B$ or $B \subseteq A$. However, this contradicts the fact that they are both in $\mathcal{A}$. $\square$

Definition 2.10. A chain $\mathcal{C} \subseteq 2^{[n]}$ is *symmetric* if, for some $0 \leq k \leq n/2$, the collection $\mathcal{C}$ contains a set of every cardinality $i \in \{k, k+1, \ldots, n-k-1, n-k\}$.

In other words, a symmetric chain is a collection $\mathcal{C} = \{X_k, X_{k+1}, \ldots, X_{n-k}\}$ of subsets of $[n]$ such that $|X_k| = k$ and, for each $k+1 \leq i \leq n-k$, the set X_i has the form $X_{i-1} \cup \{x\}$ where x is an element of $[n] \setminus X_{i-1}$.

Example 2.11. The family in Example 2.8 is not a symmetric chain in $2^{[5]}$ as it contains the empty set, which has cardinality zero, but it does not contain a set of cardinality three or five. The following is an example of a symmetric chain, when viewed in $2^{[5]}$:

$$\{\{2\}, \{2,5\}, \{1,2,5\}, \{1,2,4,5\}\}.$$

Definition 2.12. A *symmetric chain decomposition* of $2^{[n]}$ is a partition of $2^{[n]}$ into symmetric chains.

Sperner's Theorem will follow, almost immediately, from the following lemma.

Lemma 2.13. *$2^{[n]}$ has a symmetric chain decomposition.*

[3]However, note that the statement "$\mathcal{F}$ is a an antichain" is not equivalent to the statement "$\mathcal{F}$ is not a chain." There are many set systems which are neither chains nor antichains. Also, if $|\mathcal{F}| \leq 1$, then $\mathcal{F}$ is both a chain and an antichain.

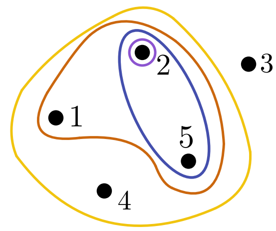


Figure 2.4: The symmetric chain in Example 2.11.

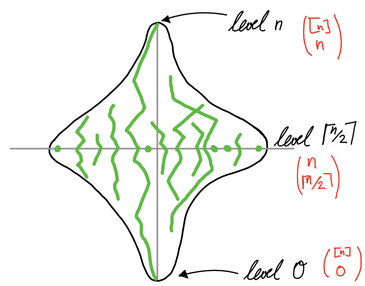


Figure 2.5: An "artist's rendition" of a symmetric chain decomposition. The (roughly) diamond-shaped blob is supposed to be $2^{[n]}$ and each of the dark green wiggly lines is supposed to be a symmetric chain. Some of the symmetric chains only have one element and are drawn as single points.

Proof. We proceed by induction on n. In the base case $n = 1$, $2^{[n]}$ itself is a symmetric chain.

Now, suppose that $n \geq 2$ and, using the inductive hypothesis, let $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_N$ be a collection of symmetric chains in $2^{[n-1]}$ which partition it. For each m such that $1 \leq m \leq N$, let

$$k_m := \min\{|A| : A \in \mathcal{C}_m\}$$

and write $\mathcal{C}_m = \{X_m^i : k_m \leq i \leq n - 1 - k_m\}$ where

$$X_m^{k_m} \subseteq X_m^{k_m+1} \subseteq \cdots \subseteq X_m^{n-1-k_m}$$

and

$$|X_m^i| = i.$$

For each $1 \leq m \leq N$, define

$$\mathcal{C}'_m := \{X_m^{k_m}, X_m^{k_m+1}, \ldots, X_m^{n-1-k_m}, X_m^{n-1-k_m} \cup \{n\}\} \text{ and}$$

$$\mathcal{C}''_m := \{X_m^{k_m} \cup \{n\}, X_m^{k_m+1} \cup \{n\}, \ldots, X_m^{n-2-k_m} \cup \{n\}\}.$$

Each of the collections $\mathcal{C}'_m$ and $\mathcal{C}''_m$ is a symmetric chain in $2^{[n]}$ and every element of $2^{[n]}$ is contained in exactly one of $\mathcal{C}'_1, \ldots, \mathcal{C}'_N, \mathcal{C}''_1, \ldots, \mathcal{C}''_N$ (see Exercise 2.2). This completes the proof. $\square$

We now present a proof of Sperner's Theorem.

Proof of Theorem 2.5. Using Lemma 2.13, we let $\mathcal{C}_1, \ldots, \mathcal{C}_N$ be a symmetric chain decomposition of $2^{[n]}$.

By definition of symmetric chain, each collection $\mathcal{C}_m$ for $1 \leq m \leq N$ must contain an element of $\binom{[n]}{\lfloor n/2 \rfloor}$. By Observation 2.9 and the fact that $\binom{[n]}{\lfloor n/2 \rfloor}$ is an antichain, it can contain at most one such set. Therefore, it contains exactly one. Since $\mathcal{C}_1, \ldots, \mathcal{C}_N$ is a partitioning of $2^{[n]}$ in which every such set contains a unique element of $\binom{[n]}{\lfloor n/2 \rfloor}$, it must hold that

$$N = \binom{n}{\lfloor n/2 \rfloor}.$$

Now, letting $\mathcal{A}$ be any antichain in $2^{[n]}$, since $\mathcal{C}_1, \ldots, \mathcal{C}_N$ partitions $2^{[n]}$, we have

$$|\mathcal{A}| = \sum_{m=1}^{N} |\mathcal{A} \cap \mathcal{C}_m|.$$

By Observation 2.9, each of the summands on the right side of this equation is at most one, and so $|\mathcal{A}| \leq N = \binom{n}{\lfloor n/2 \rfloor}$, as desired. □

2.2 The LYM Inequalities

Figure 2.6: "Chain."

We now explore some strengthenings of Sperner's Theorem known as the LYM Inequality[4] and the Local LYM Inequality; first, a definition.

Definition 2.14. A chain $\mathcal{C} \subseteq 2^{[n]}$ is said to be *maximal* if there does not exist a chain $\mathcal{C}'$ such that

$$\mathcal{C} \subsetneq \mathcal{C}' \subseteq 2^{[n]}.$$

[4]There is an exercise in Math 322 at UVic which leads through a proof of the LYM Inequality [192, Chapter 3, Exercise 8].

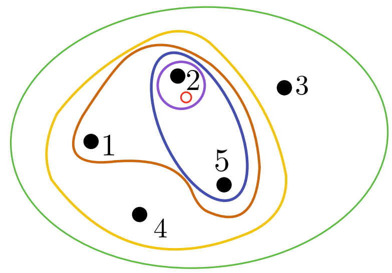


Figure 2.7: A maximal chain in $2^{[5]}$, where the red set is $\emptyset$.

Since each chain can contain at most one element of $\binom{[n]}{k}$ for each k (by Example 2.3 and Observation 2.9), we can equivalently say that a chain is maximal if it has $n+1$ elements, one of each cardinality between 0 and n.

Theorem 2.15 (The LYM Inequality [30, 190, 195, 267])**.** *If $\mathcal{A} \subseteq 2^{[n]}$ is an antichain, then*

$$\sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|A|}} \leq 1.$$

Proof. Let $\mathcal{C}$ be a maximal chain in $2^{[n]}$ selected uniformly at random among all such chains. For each $A \in 2^{[n]}$, let I_A be the indicator function of the event that $A \in \mathcal{C}$; that is,

$$I_A = \begin{cases} 1 & \text{if } A \in \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{A}$ be an antichain. For any choice of $\mathcal{C}$, we have

$$\sum_{A \in \mathcal{A}} I_A = |\mathcal{A} \cap \mathcal{C}| \leq 1$$

by Observation 2.9. Therefore, if we take the expectation of both sides, then we get, by linearity of expectation (see Appendix B),

$$1 \geq \mathbb{E}\left(\sum_{A \in \mathcal{A}} I_A\right) = \sum_{A \in \mathcal{A}} \mathbb{E}(I_A) = \sum_{A \in \mathcal{A}} \mathbb{P}(A \in \mathcal{C}).$$

Now, for any $0 \leq k \leq n$, the chain $\mathcal{C}$ contains exactly one element of $\binom{[n]}{k}$, and any such set is equally likely to be contained in $\mathcal{C}$, by symmetry. Therefore, for any $A \subseteq [n]$, it holds that

$$\mathbb{P}(A \in \mathcal{C}) = \frac{1}{\binom{n}{|A|}}.$$

Putting these inequalities together completes the proof. $\square$

Observe that the LYM Inequality is tight when $\mathcal{A} = \binom{[n]}{k}$ for some k. One way to think about the LYM Inequality is to view each subset A of $[n]$ as being associated to a "cost" of $\binom{n}{|A|}^{-1}$. The

LYM Inequality says that, for any antichain, the total cost of the elements within it must be at most one. Note that, since $\binom{n}{k}$ is maximized when k is $\lfloor n/2 \rfloor$, the sets of lowest cost are those of size $\lfloor n/2 \rfloor$. From this, it is easily observed that no antichain can have more than $\binom{n}{\lfloor n/2 \rfloor}$ elements, and so the LYM Inequality implies Sperner's Theorem.

Sperner's Theorem can be equivalently thought of as a bound on the size of a family which does not contain a chain of cardinality two. In the same way as above, the LYM Inequality can also be used to determine the maximum size of a family in $2^{[n]}$ which does not contain a larger chain.

Corollary 2.16 (Erdős [91])**.** *For $2 \leq k \leq n$, if $\mathcal{F} \subseteq 2^{[n]}$ such that $\mathcal{F}$ does not contain a chain with k elements, then $|\mathcal{F}|$ is bounded above by the sum of the $k-1$ largest binomial coefficients $\binom{n}{m}$.*

Proof. Exercise 2.16. □

Next, we apply the LYM Inequality to obtain a result on the "expansion" of a family of sets on the kth layer of $2^{[n]}$, known as the Local LYM Inequality; first, a definition.

Definition 2.17. Given a family $\mathcal{F} \subseteq \binom{[n]}{k}$, the *shadow* of $\mathcal{F}$ is defined to be

$$\partial\mathcal{F} := \{B : |B| = k-1 \text{ and there exists } A \in \mathcal{F} \text{ such that } B \subseteq A\}.$$

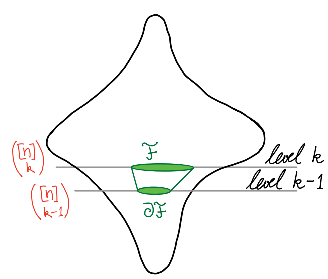


Figure 2.8: A family $\mathcal{F} \subseteq \binom{[n]}{k}$ and its shadow.

Corollary 2.18 (The Local LYM Inequality)**.** *For $1 \leq k \leq n$, if $\mathcal{F} \subseteq \binom{[n]}{k}$, then*

$$\frac{|\partial\mathcal{F}|}{\binom{n}{k-1}} \geq \frac{|\mathcal{F}|}{\binom{n}{k}}.$$

Proof. Define

$$\mathcal{A} := \mathcal{F} \cup \left(\binom{[n]}{k-1} \setminus \partial\mathcal{F}\right).$$

Then $\mathcal{A}$ is clearly an antichain (do you see why?). Therefore, by the LYM Inequality,

$$\sum_{A \in \mathcal{A}} \frac{1}{\binom{n}{|A|}} = \frac{|\mathcal{F}|}{\binom{n}{k}} + \frac{\binom{n}{k-1} - |\partial\mathcal{F}|}{\binom{n}{k-1}} \leq 1$$

and we are done after a bit of rearranging. □

Figure 2.9: "Shadows."

2.3 The Kruskal–Katona Theorem

As we saw in the previous section, the Local LYM Inequality provides a lower bound on the cardinality of the shadow of a family $\mathcal{F} \subseteq \binom{[n]}{k}$ in terms of the cardinality of $\mathcal{F}$. However, this lower bound is rarely optimal. Therefore, it is natural to wonder the following: Given $1 \leq k \leq n$ and $1 \leq m \leq \binom{n}{k}$, how small can the shadow of a collection $\mathcal{F} \subseteq \binom{[n]}{k}$ with $|\mathcal{F}| = m$ be?

Example 2.19. As a warm-up, consider the case $k = 2$; that is, $\mathcal{F} \subseteq \binom{\mathbb{N}}{2}$. The shadow of $\mathcal{F}$ is the collection of all sets of size one contained in an element of $\mathcal{F}$. Therefore, in this special case,

$$|\partial\mathcal{F}| = \left|\bigcup_{A\in\mathcal{F}} A\right|.$$

So, the smallest shadow is obtained by any collection of sets of size two which covers as few points as possible. For $m \geq 1$, if we let ℓ be the unique integer so that

$$\binom{\ell-1}{2} < m \leq \binom{\ell}{2},$$

then, if you think about it, if $|\mathcal{F}| = m$, then the the union of the sets in $\mathcal{F}$ must be at least ℓ, and this can be achieved by taking $\mathcal{F} \subseteq \binom{[\ell]}{2}$ with $|\mathcal{F}| = m$.

Example 2.19 suggests the following.

Intuition. If you want $\partial\mathcal{F}$ to be small, then it is wise to keep the union of sets in $\mathcal{F}$ small. That is, one should avoid using many elements.

Let's play with sets of cardinality three.

Example 2.20. Let us think about how we would construct a family $\mathcal{F} \subseteq \binom{\mathbb{N}}{3}$ with $|\mathcal{F}| = m$ for small values of t which minimizes the cardinality of $\partial\mathcal{F}$. For $m = 1$, the family $\mathcal{F}$ consists of a single set of size three, and so the cardinality of the shadow is always equal to $\binom{3}{2}$.

Things start getting more interesting for $m = 2$. Without loss of generality, assume that $\mathcal{F}$ contains the set $\{1,2,3\}$. If the other element of $\mathcal{F}$ is $\{a,b,c\}$, then

$$\partial\mathcal{F} = \{\{1,2\},\{1,3\},\{2,3\}\} \cup \{\{a,b\},\{a,c\},\{b,c\}\}.$$

Therefore, if we are trying to make $\partial\mathcal{F}$ as small as possible, we want the "overlap" between the collections $\{\{1,2\},\{1,3\},\{2,3\}\}$ and $\{\{a,b\},\{a,c\},\{b,c\}\}$ to be as large as possible. Clearly, this occurs when $\{1,2,3\}$ and $\{a,b,c\}$ have as many elements in common as possible and so it is optimal to take $\{a,b,c\} = \{1,2,4\}$. Note that, of course, this choice is not unique; e.g., $\{a,b,c\} = \{2,3,7\}$ is also optimal. Note that this matches our earlier intuition that it is smart to avoid introducing new elements of $\mathbb{N}$ into the union of sets in $\mathcal{F}$ as much as possible.

Now, if we apply similar intuition to the case $m = 3$, it seems to be a good idea to take three sets which have large pairwise intersections. For the sake of discussion, suppose that $\mathcal{F}$ contains the sets $\{1,2,3\},\{1,2,4\}$ and a third set $\{a,b,c\}$. Then

$$\partial\mathcal{F} = \{\{1,2\},\{1,3\},\{2,3\},\{1,4\},\{2,4\}\} \cup \{\{a,b\},\{a,c\},\{b,c\}\}.$$

In particular, $\partial\mathcal{F}$ already contains all elements of $\binom{[4]}{2}$, with the exception of $\{3,4\}$. So, if we take $\{a,b,c\}$ to be a subset of $\{1,2,3,4\}$, then, at worst, the shadow will have cardinality 6. On the other hand, if $\{a,b,c\}$ contains an element of $\{5,6,7\}$, then at least two of the sets $\{a,b\},\{a,c\}$ and $\{b,c\}$ will not be contained in $\binom{[4]}{2}$, and so $\partial\mathcal{F}$ will contain at least 7 elements. With a bit more thinking, these arguments lead us to the conclusion that an optimal choice for $\mathcal{F}$ is $\{\{1,2,3\},\{1,2,4\},\{1,3,4\}\}$. But also, taking any three sets from $\binom{[4]}{3}$ would be equally good.

Given the answer for $m = 3$, the case $m = 4$ actually becomes trivial. If we let $\mathcal{F} = \binom{[4]}{3}$, then the shadow again has cardinality 6, which must be optimal, as it cannot be better than the best example in the case $m = 3$.

Now, when $m = 5$, the best thing to do is to take $\mathcal{F} = \binom{[4]}{3} \cup \{\{1,2,5\}\}$, which has a shadow of cardinality 8. For $m = 6$, the best is $\mathcal{F} = \binom{[4]}{3} \cup \{\{1,2,5\},\{1,3,5\}\}$ and for $m = 7$ it is $\mathcal{F} = \binom{[4]}{3} \cup \{\{1,2,5\},\{1,3,5\},\{2,3,5\}\}$.

But why, you may ask? It is a bit tedious to continue to rule out all other possibilities as we have been doing, but the vague intuition is as follows. When $\binom{4}{3} < t \leq \binom{5}{3}$, we have no choice but to introduce a fifth element of $[n]$, say, the number 5, into the union of sets in $\mathcal{F}$. Assuming that $\mathcal{F}$ contains all of the sets in $\binom{[4]}{3}$ (which we haven't justified, but let's go with it), then the "new" sets that we get in the shadow are of the form $\{a,5\}$ where $\mathcal{F}$ contains a set of the form $\{a,b,5\}$. So, if we let $\mathcal{F}' = \{\{a,b\} : \{a,b,5\} \in \mathcal{F}\}$, then it seems reasonable that, to make the shadow of $\mathcal{F}$ small, we would want to minimize the shadow of $\mathcal{F}'$; however, we already learned how to do that in Example 2.19.

Example 2.20 now suggests:

Intuition. When you are forced to introduce a "new" element into the union of the sets in $\mathcal{F}$, it is wise to keep the union of the sets of $\mathcal{F}$ which contain that element small.

Still, it seems that it is not obvious what the best collection is for minimizing the shadow. One thing that complicates things is that, as we saw in the above examples, the optimal construction usually isn't unique. However, the intuition that we should "avoid introducing new elements as long as possible" does point us towards a certain construction. Consider the following.

Definition 2.21. Given distinct subsets $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_k\}$ of $\mathbb{N}$ with $a_1 < \cdots < a_k$ and $b_1 < \cdots < b_k$, we write

$$A \prec_{\text{colex}} B$$

to mean that $a_j < b_j$ where $j = \max\{i : a_i \neq b_i\}$. This ordering is called the *colexicographic order* (or *colex order*, for short) on $\binom{\mathbb{N}}{k}$.

In other words, $A \prec_{\text{colex}} B$ if and only if there exists m such that the mth largest element of A is less than the mth largest element of B and all of the larger elements of A and B coincide.

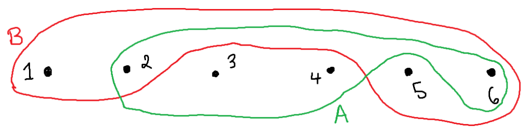


Figure 2.10: If $A = \{2, 3, 4, 6\}$ is the green set and $B = \{1, 2, 5, 6\}$ is the red set, then $A \prec_{\text{colex}} B$ because the largest elements of A and B are the same, but the second largest element of B (which is 5) is larger than the second largest element of A (which is 4). That is, $a_j < b_j$ where j is the latest index on which A and B differ.

Example 2.22. The first seven elements of $\binom{\mathbb{N}}{2}$ under the colex order are

$$\{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 4\}, \{2, 4\}, \{3, 4\}, \{1, 5\}.$$

The colex order seems to have the properties that we described above; that is, the colex order aims to avoid introducing large numbers as long as possible. It also has a nice "recursive" structure in the sense that, when a new number, say ℓ, is introduced, if look at the sets of size $k - 1$ that you get from deleting ℓ from all sets in which ℓ is the maximum element, then they are ordered according to the colex order on $\binom{\mathbb{N}}{k-1}$. Just for completeness, let us also mention the more familiar lexicographic order. We won't actually use it for anything, though.

Definition 2.23. Given distinct subsets $A = \{a_1, \ldots, a_k\}$ and $B = \{b_1, \ldots, b_k\}$ of $\mathbb{N}$ with $a_1 < \cdots < a_k$ and $b_1 < \cdots < b_k$, we write

$$A \prec_{\text{lex}} B$$

to mean that $a_j < b_j$ where $j = \min\{i : a_i \neq b_i\}$. This ordering is called the *lexicographic order* (or *lex order*, for short) on $\binom{\mathbb{N}}{k}$.

Example 2.24. If A and B are the sets from Figure 2.3, then $B \prec_{\text{lex}} A$ because the smallest element of B is 1, which is less than the smallest element of A which is 2.

Example 2.25. The first seven elements of $\binom{\mathbb{N}}{2}$ under lex order are

$$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{1, 5\}, \{1, 6\}, \{1, 7\}, \{1, 8\}.$$

Our focus in this section is on proving the following theorem, which confirms our intuition that colex order is useful for minimizing the size of the shadow.

Theorem 2.26 (The Kruskal–Katona Theorem [144, 166])**.** *For $k \ge 1$ and $m \ge 1$, let $X_1, \dots, X_m$ be the first m sets of $\binom{\mathbb{N}}{k}$ under colex order. Then every $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ with $|\mathcal{F}| = m$ satisfies*

$$|\partial \mathcal{F}| \ge |\partial\{X_1, \dots, X_m\}|.$$

In other words, the Kruskal–Katona Theorem says that "initial segments" of the colex order have the smallest shadow. An important part of the proof is to understand how colex order works.

Example 2.27. Given a set $A \in \binom{\mathbb{N}}{k}$, how do we determine the position of A in the colex order? For example, consider $A = \{3, 7, 9\}$. Then every set $B \in \binom{[8]}{3}$ comes before A. Also, every set B of size 3 whose largest element is 9 and whose second largest element is at most 6 comes before A. Finally, sets whose largest and second largest elements are 9 and 7 whose smallest element is at most 2 come before A. These three disjoint classes actually describe all of the sets that come before A. So, the number of sets that come before A is

$$\binom{8}{3} + \binom{6}{2} + \binom{2}{1}.$$

The argument in this example generalizes straightforwardly to the following.

Lemma 2.28. *If $A = \{a_1, \dots, a_k\}$ where $a_1 < a_2 < \cdots < a_k$, then A is the mth set of $\binom{\mathbb{N}}{k}$ under the colex ordering where*

$$m = 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i}.$$

Here is a simple, but surprisingly useful, corollary of Lemma 2.28.

Corollary 2.29. *Let $k \ge 1$. Then, for every integer $m \ge 1$, there is a unique increasing sequence $a_1, \dots, a_k$ of positive integers such that*

$$m = 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i}.$$

Proof. For existence, let A be the mth element of $\binom{\mathbb{N}}{k}$ under colex. Let $a_1, \dots, a_k$ be the elements of A in increasing order. Apply Lemma 2.28. Uniqueness follows from the fact that colex is a total order. □

Before trying to prove the Kruskal–Katona Theorem, it is useful to get a better understanding of what our "target" lower bound is. That is, how large is the the shadow of the first m elements under colex?

Corollary 2.30. *Let $m \ge 1$ and let $a_1, \dots, a_k$ be an increasing sequence of positive integers such that*

$$m = 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i}.$$

Let $X_1, \dots, X_m$ be the first m elements of $\binom{\mathbb{N}}{k}$ under colex. Then

$$|\partial\{X_1, \dots, X_m\}| = \sum_{i=1}^{k} \binom{a_i - 1}{i - 1}.$$

Proof. Let $\mathcal{I}_m = \{X_1, \ldots, X_m\}$. By Lemma 2.28, we have

$$X_m = \{a_1, a_2, \ldots, a_k\}.$$

Therefore, all sets of size k whose largest element is less than a_k are in $\mathcal{I}_m$. This implies that all sets of size $k-1$ whose largest element is less than a_k are in $\partial\mathcal{I}_m$. The number of such sets is $\binom{a_k-1}{k-1}$. Also, all sets of size k whose largest element is a_k and whose second largest element is less than a_{k-1} are in $\mathcal{I}_m$. This implies that all sets of size $k-1$ whose largest element is a_k and second largest is less than a_{k-1} are in $\partial\mathcal{I}_m$. The number of such sets is $\binom{a_{k-1}-1}{k-2}$. Continuing this way, we eventually get the result. $\square$

This allows us to restate the Kruskal–Katona Theorem as follows: *For $k \geq 1$ and $m \geq 1$, let $a_1, \ldots, a_k$ be a sequence of positive integers such that*

$$m = 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i}.$$

Then every $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ with $|\mathcal{F}| = m$ satisfies

$$|\partial\mathcal{F}| \geq \sum_{i=1}^{k} \binom{a_i - 1}{i-1}.$$

A natural strategy to prove the Kruskal–Katona Theorem is to start with a collection $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ and try to replace some of the sets in $\mathcal{F}$ with sets that come earlier under the colex order. The natural way to modify a set A to make it come earlier under colex is to replace a large element $j \in A$ with a small element $i \notin A$. The simplest operation for doing this is as follows:

Definition 2.31. For distinct $i, j \in \mathbb{N}$, let $C_{i,j} : 2^{\mathbb{N}} \to 2^{\mathbb{N}}$ be defined by

$$C_{i,j}(A) := \begin{cases} (A \setminus \{j\}) \cup \{i\} & \text{if } j \in A \text{ and } i \notin A, \\ A & \text{otherwise} \end{cases}$$

for each $A \subseteq \mathbb{N}$. We call $C_{i,j}(A)$ the *(i,j)-compression* of A.

The function $C_{i,j}$ is certainly not a bijection. If we let

$$\mathcal{S}_{i,j} := \{A \subseteq \mathbb{N} : i \notin A \text{ and } j \in A\},$$

then $C_{i,j}$ maps $\mathcal{S}_{i,j}$ bijectively to $\mathcal{S}_{j,i}$, but it does not map any set into $\mathcal{S}_{i,j}$. See Figure 2.3. Since $C_{i,j}$ is not a bijection, simply replacing each set A in a collection $\mathcal{F}$ with $C_{i,j}(A)$ can change the cardinality of $\mathcal{F}$. A smarter way to apply $C_{i,j}$ to a collection of sets is as follows.

Definition 2.32. For distinct $1 \leq i, j \leq n$ and $\mathcal{F} \subseteq 2^{[n]}$, define

$$C_{i,j}(\mathcal{F}) := \{C_{i,j}(A) : A \in \mathcal{F}\} \cup \{A \in \mathcal{F} : C_{i,j}(A) \in \mathcal{F}\}.$$

Call $C_{i,j}(\mathcal{F})$ the *(i,j)-compression of $\mathcal{F}$*.

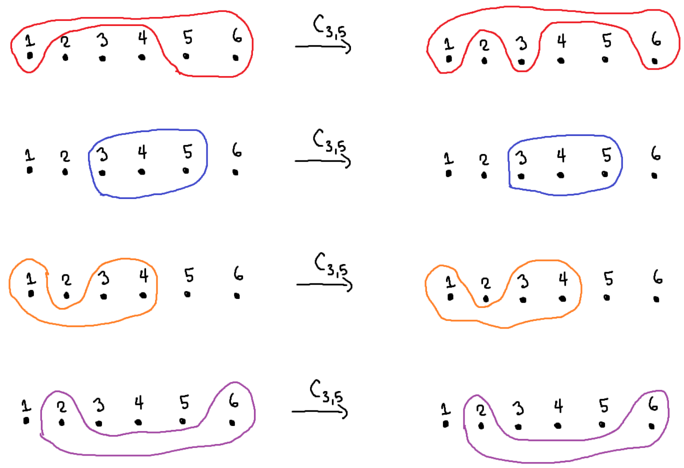


Figure 2.11: The set A and the set $C_{3,5}(A)$ for four different subsets of $[6]$. Only the first one satisfies $C_{3,5}(A) \neq A$ because $5 \in A$ and $3 \notin A$.

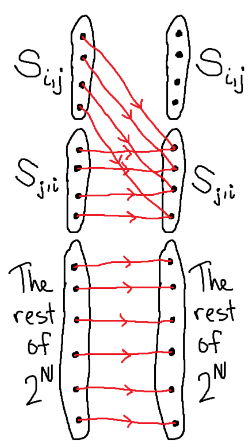


Figure 2.12: An illustration of how the function $C_{i,j}$ behaves on $2^{\mathbb{N}}$.

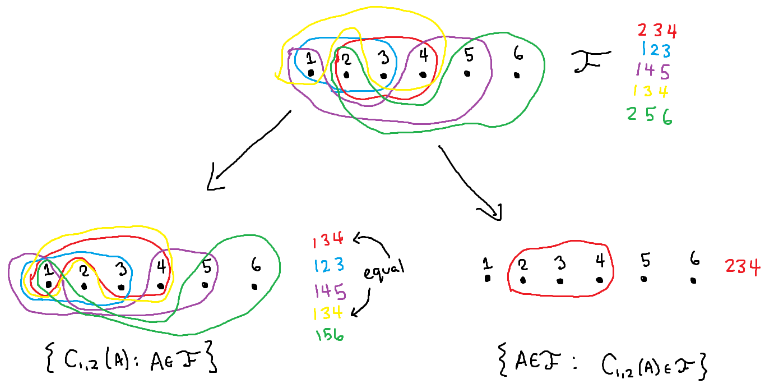


Figure 2.13: The compression $C_{1,2}$ applied to a family $\mathcal{F} \subseteq \binom{\mathbb{N}}{3}$.

The proof of the following lemma can be more or less understood by staring at Figure 2.3 long enough. We will still give a proper proof, though.

Lemma 2.33. *For any $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ and distinct $i, j \geq 1$,*

$$|C_{i,j}(\mathcal{F})| = |\mathcal{F}|.$$

Proof. We describe an explicit bijection $\varphi : \mathcal{F} \to C_{i,j}(\mathcal{F})$. For $A \in \mathcal{F}$, let

$$\varphi(A) = \begin{cases} A & \text{if } A \in C_{i,j}(\mathcal{F}) \\ C_{i,j}(A) & \text{otherwise.} \end{cases}$$

Let's argue that φ is surjective. Any set B in $C_{i,j}(\mathcal{F})$ which is also in $\mathcal{F}$ satisfies $\varphi(B) = B$, and so it is covered. If $B \in C_{i,j}(\mathcal{F})$ and it is not in $\mathcal{F}$, then the set $A = C_{j,i}(B)$ must be in $\mathcal{F}$. We claim that $\varphi(A) = B$. If not, then it must be the case that $\varphi(A) = A$, which only happens if $A \in C_{i,j}(\mathcal{F})$. But the only reason that A could be in $C_{i,j}(\mathcal{F})$ is if $B \in \mathcal{F}$, which we already said is not the case; contradiction.

Now, we verify that it is injective. Suppose $A_1, A_2 \in \mathcal{F}$ and $\varphi(A_1) = \varphi(A_2)$. If neither of A_1 nor A_2 are in $C_{i,j}(\mathcal{F})$, then we get $A_1 = \varphi(A_1) = \varphi(A_2) = A_2$. If $A_1 \in C_{i,j}(\mathcal{F})$ and $A_2 \notin C_{i,j}(\mathcal{F})$, then we have $A_1 = \varphi(A_1) = \varphi(A_2) = C_{i,j}(A_2)$. But then $C_{i,j}(A_2) \in \mathcal{F}$ which, by definition of $C_{i,j}(\mathcal{F})$, implies that $A_2 \in \mathcal{F}$, which we said was not the case; contradiction. If neither A_1 nor A_2 are in $C_{i,j}(\mathcal{F})$, then $C_{i,j}(A_1) = \varphi(A_1) = \varphi(A_2) = C_{i,j}(A_2)$. The only way for $C_{i,j}(A_1)$ to equal $C_{i,j}(A_2)$ while $A_1, A_2 \notin C_{i,j}(\mathcal{F})$ is if $A_1 = A_2$. This completes the proof. $\square$

What is less clear, and more difficult to visualize, is that applying $C_{i,j}$ to a collection $\mathcal{F}$ does not increase the size of the shadow.

Lemma 2.34. *For any $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ and distinct $i, j \geq 1$,*

$$|\partial C_{i,j}(\mathcal{F})| \leq |\partial\mathcal{F}|.$$

Proof. What we will actually prove is that

$$\partial C_{i,j}(\mathcal{F}) \subseteq C_{i,j}(\partial\mathcal{F}). \tag{2.35}$$

This will be sufficient since, by Lemma 2.33, $|C_{i,j}(\partial\mathcal{F})| = |\partial\mathcal{F}|$.

So, we need to show that every set $B \in \partial C_{i,j}(\mathcal{F})$ is in $C_{i,j}(\partial\mathcal{F})$. None of the individual steps in proving this are hard. The difficulty mainly comes from digesting what it "means" and then dealing the multitude of cases that arise.

Let's think this through. The first thing to ask ourselves is "where does a set in $\partial C_{i,j}(\mathcal{F})$ come from?" There are two different answers to this. An element of this collection can either be of the form

$$C_{i,j}(A) \setminus \{\ell\} \qquad \text{for some } A \in \mathcal{F} \text{ and } \ell \in C_{i,j}(A), \text{ or} \tag{2.36}$$

$$A \setminus \{\ell\} \qquad \text{for some } A \in \mathcal{F} \text{ such that } C_{i,j}(A) \in \mathcal{F} \text{ and } \ell \in A. \tag{2.37}$$

Whichever of these two cases we are in, if $\ell \notin \{i,j\}$, then it is fairly straightforward to prove what we want to prove. The point is that deleting the element ℓ is sort of "irrelevant" from the point of view of the (i,j)-compression. Still, let's go through the argument. In case (2.36), the fact that $\ell \in C_{i,j}(A)$ and $\ell \neq i$ tells us that $\ell \in A$ (by definition of $C_{i,j}$). Thus, $A \setminus \{\ell\} \in \partial\mathcal{F}$. Since $\ell \notin \{i,j\}$, we have $A \cap \{i,j\} = (A \setminus \{\ell\}) \cap \{i,j\}$, and so

$$C_{i,j}(A) \setminus \{\ell\} = C_{i,j}(A \setminus \{\ell\}) \in C_{i,j}(\partial\mathcal{F}).$$

Now, in case (2.37), note that $A \in \mathcal{F}$ and $C_{i,j}(A) \in \mathcal{F}$ implies that $A \setminus \{\ell\} \in \partial\mathcal{F}$ and $C_{i,j}(A \setminus \{\ell\}) \in \partial\mathcal{F}$. This implies that $A \setminus \{\ell\} \in C_{i,j}(\partial\mathcal{F})$.

So, what remains are the cases $\ell = i$ and $\ell = j$. Let's start with $\ell = j$. In case (2.36), we have $j \in C_{i,j}(A)$ which implies that $C_{i,j}(A) = A$. Therefore,

$$C_{i,j}(A) \setminus \{j\} = A \setminus \{j\} = C_{i,j}(A \setminus \{j\}) \in C_{i,j}(\partial\mathcal{F})$$

by Definition 2.31. Case (2.37) works by the same proof; that is, the expression written above still holds (and we don't even need the first equality).

Finally, let's think about $\ell = i$. We again consider the cases (2.36) and (2.37) separately. Suppose first that $A \in \mathcal{F}$ and $i \in C_{i,j}(A)$. If $j \notin C_{i,j}(A)$, then A contains at most one of i or j, which implies that $A \setminus \{i,j\} \in \partial\mathcal{F}$, and we get

$$C_{i,j}(A) \setminus \{i,j\} = A \setminus \{i,j\} = C_{i,j}(A \setminus \{i,j\}) \in C_{i,j}(\partial\mathcal{F}).$$

On the other hand, if $j \in C_{i,j}(A)$, then both $A \setminus \{i\}$ and $A \setminus \{j\}$ are in $\partial\mathcal{F}$. Note that $A \setminus \{j\} = C_{i,j}(A \setminus \{i\})$ and so

$$C_{i,j}(A) \setminus \{i\} = A \setminus \{i\} \in C_{i,j}(\partial\mathcal{F})$$

by Definition 2.32 and the fact that $A \setminus \{j\} \in \partial\mathcal{F}$. Now, let's think about case (2.37). Actually, it turns out that there is nothing to do here. That is, in the second case, we are dealing with $A \setminus \{i\}$ for a set A such that $i \in A$. But if $i \in A$, then $C_{i,j}(A) = A$, and so, actually, we are done by the first case. $\square$

The notion of compression leads us to a natural strategy for trying to prove the Kruskal–Katona Theorem. That is, start with a finite collection $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$. While there exists $i < j$ such that $C_{i,j}(\mathcal{F}) \neq \mathcal{F}$, replace $\mathcal{F}$ with $C_{i,j}(\mathcal{F})$. By Lemma 2.33, this does not change the size of the collection and, by Lemma 2.34, this does not increase the size of the shadow. Imagine that we keep repeating this until no such i and j exist. Two natural questions arise:

1. Does this eventually terminate?
2. Does it eventually reach an initial segment of colex?

The answer to the first question here is "yes." One way to see this is by defining a weight function. For a finite set $A \subseteq \mathbb{N}$, define the *weight* of A to be

$$w(A) := \sum_{i \in A} 2^i.$$

Then define the *weight* of a finite collection $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ to be

$$w(\mathcal{F}) = \sum_{A \in \mathcal{F}} w(A).$$

The key observation is that $C_{i,j}$ does not increase the weight of a set and, if $C_{i,j}(A) \neq A$, then the weight of $C_{i,j}(A)$ is strictly less than that of A. Since the weight of a collection is an integer, and applying $C_{i,j}$ for $i < j$ can only decrease the weight of a collection, we see that the process described above eventually terminates.

Thus, to prove the Kruskal–Katona Theorem, it suffices to consider a collection $\mathcal{F}$ such that $C_{i,j}(\mathcal{F}) = \mathcal{F}$ for all $i < j$. The following definition describes this situation.

Definition 2.38. For distinct $i, j \geq 1$, say that $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ is (i,j)-*compressed* if $C_{i,j}(\mathcal{F}) = \mathcal{F}$. Say that $\mathcal{F}$ is *compressed* if it is (i,j)-compressed for all $i < j$.

Now, what about that second question? If $\mathcal{F}$ is compressed, does it follow that $\mathcal{F}$ is an initial segment of colex? Sadly, it does not, as the next example shows.

Example 2.39. Consider

$$\mathcal{F} = \{\{1,2,3\}, \{1,2,4\}, \{1,2,5\}, \{1,2,6\}, \{1,2,7\}\}.$$

Clearly, this is not an initial segment of colex. However, if you think about it, you can see that $\mathcal{F}$ is compressed.

Indeed, if $i \in \{1,2\}$, then $C_{i,j}(A) = A$ for all $j > i$ and $A \in \mathcal{F}$. Also, if $3 \leq i < j \leq 7$ and $A \in \mathcal{F}$, then either $C_{i,j}(A) = A$ or $C_{i,j}(A) \in \mathcal{F}$. Thus, $C_{i,j}(\mathcal{F}) = \mathcal{F}$.

At this point, it may feel like our strategy of trading large elements for small ones has failed. However, the situation is not as bad as it seems. There are several directions that we can go. One way is to consider more powerful types of compressions in which, instead of trading one element for another, we trade a whole set of elements for another set that comes earlier in colex. If you are sufficiently careful in how you do it, then this can work; this is the strategy followed in [169], for example. We won't do it this way, but it is worth being aware that it is possible. Instead, we will follow an inductive argument.

Proof of Theorem 2.26. We proceed by induction on $k + m$. As a base case, suppose that $k = 1$. This case is easy, as every collection $\mathcal{F} \subseteq \binom{\mathbb{N}}{1}$ has the same shadow; namely, $\{\emptyset\}$. In what follows, we assume that $k \geq 2$ and let $m \geq 1$. By Corollary 2.30, we can let $a_1, \ldots, a_k$ be a sequence of positive integers such that

$$m = 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i}. \tag{2.40}$$

Let $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ such that $|\mathcal{F}| = m$. By Lemmas 2.33 and 2.34, we can assume that $\mathcal{F}$ is compressed. Define

$$\mathcal{F}_1 := \{A \in \mathcal{F} : 1 \in A\}, \text{ and}$$
$$\mathcal{F}_0 := \mathcal{F} \setminus \mathcal{F}_1.$$

Here's a cool thing about $\mathcal{F}$ being compressed. Imagine that we start with a set $A \in \mathcal{F}_0$. Take an element ℓ of A and delete it; this gives us an element of $\partial\mathcal{F}_0$. Now, add the element 1 to the resulting set. This process yields $C_{1,\ell}(A)$, which must be an element of $\mathcal{F}_1$ because $\mathcal{F}$ is compressed. The last step of this process (adding 1 to a set in $\partial\mathcal{F}_0$) is an injection from $\partial\mathcal{F}_0$ to $\mathcal{F}_1$. Therefore, we have the following

$$|\mathcal{F}_1| \geq |\partial\mathcal{F}_0|. \tag{2.41}$$

The above equation bounds the size of $\mathcal{F}_1$ relative to the shadow of $\mathcal{F}_0$. Let's turn this into an absolute bound.

Claim 2.42. $|\mathcal{F}_1| \geq 1 + \sum_{i=2}^{k} \binom{a_i-2}{i-1}$.

Proof of Claim 2.42. We suppose that it is false and derive a contradiction. By (2.40), we get

$$|\mathcal{F}_0| = |\mathcal{F}| - |\mathcal{F}_1| = 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i} - |\mathcal{F}_1| \geq 1 + \sum_{i=1}^{k} \binom{a_i - 1}{i} - \sum_{i=2}^{k} \binom{a_i - 2}{i - 1}$$
$$= a_i + \sum_{i=2}^{k} \left[\binom{a_i - 1}{i} - \binom{a_i - 2}{i - 1}\right] = 1 + \binom{a_1 - 1}{1} + \sum_{i=2}^{k} \binom{a_i - 2}{i}.$$

In conclusion,

$$|\mathcal{F}_0| \geq 1 + \binom{a_1 - 1}{1} + \sum_{i=2}^{k} \binom{a_i - 2}{i}. \tag{2.43}$$

At first, this may seem pretty irrelevant for proving the claim. But here's the kicker. Knowing that $|\mathcal{F}_0|$ is large implies that its shadow is large by the induction hypothesis, which, by (2.41), will implies that $|\mathcal{F}_1|$ is large as well.

Here are the details. What we would like to do is simply say that (2.43), mixed with induction, immediately implies that

$$|\partial\mathcal{F}_0| \geq \binom{a_1 - 1}{0} + \sum_{i=2}^{k} \binom{a_i - 2}{i - 1} \geq 1 + \sum_{i=2}^{k} \binom{a_i - 2}{i - 1} \tag{2.44}$$

which proves the claim. This almost just works. However, we need to be a little bit more careful. We can't *quite* apply induction immediately here because we don't know that the sequence $a_1, a_2 - 1, \ldots, a_k - 1$ is increasing. That is, we could have $a_1 = a_2 - 1$. We can get around this issue with a little bit of tinkering. If $a_1 \geq 2$, then we can use (2.43) and the fact that that $a_1 - 1, a_2 - 1, \ldots, a_k - 1$ is increasing to derive (2.44), since $\binom{a_1-2}{0} = 1$. So, we only have a problem in the case $a_1 = 1$ and $a_2 = 2$. Define j to be the largest index such that $a_j = j$. Then, for $2 \leq i \leq j$, we have $\binom{a_i-2}{i} = 0 = \binom{a_i-1}{i}$. So, we can rewrite (2.43) as

$$|\mathcal{F}_0| \geq 1 + \binom{a_1 - 1}{1} + \sum_{i=2}^{j} \binom{a_i - 1}{i} + \sum_{i=j+1}^{k} \binom{a_i - 2}{i}.$$

The sequence $a_1, a_2, \ldots, a_j, a_{j+1}-1, \ldots, a_k-1$ is genuinely increasing, so we can derive (2.44) from this inequality and induction.

Finally, (2.41) and (2.44) directly imply the claim. $\square$

We are now in position to finish the proof off. Define

$$\mathcal{G} := \{A \setminus \{1\} : A \in \mathcal{F}_1\}.$$

Note that $\mathcal{G} \subseteq \partial\mathcal{F}$ and that $|\mathcal{G}| = |\mathcal{F}_1|$. Define a map $\psi : \mathcal{G} \cup \partial\mathcal{G} \to \partial\mathcal{F}$ by defining, for each B in the domain,

$$\psi(B) := \begin{cases} B & \text{if } B \in \mathcal{G}, \\ B \cup \{1\} & \text{if } B \in \partial\mathcal{G}. \end{cases}$$

The function ψ is injective; therefore $|\partial\mathcal{F}| \geq |\mathcal{G}| + |\partial\mathcal{G}|$. By Claim 2.42, the facts that $|\mathcal{G}| = |\mathcal{F}_1|$ and $\mathcal{G} \subseteq \binom{\mathbb{N}}{k-1}$, and induction, we have

$$|\partial\mathcal{F}| \geq |\mathcal{G}| + |\partial\mathcal{G}| \geq 1 + \sum_{i=2}^{k} \binom{a_i - 2}{i-1} + \sum_{i=2}^{k} \binom{a_i - 2}{i-2}$$

$$= 1 + \sum_{i=2}^{k} \binom{a_i - 1}{i-1} = \sum_{i=1}^{k} \binom{a_i - 1}{i-1}.$$

WOW! Right? $\square$

The following is a special case of Theorem 2.26 which has a particularly elegant formulation.

Corollary 2.45 (Lovász [183])**.** *If $1 \leq k \leq \ell$ and $\mathcal{F} \subseteq \binom{\mathbb{N}}{k}$ such that $|\mathcal{F}| = \binom{\ell}{k}$, then*

$$|\partial\mathcal{F}| \geq \binom{\ell}{k-1}.$$

Proof. The first $\binom{\ell}{k}$ elements of $\binom{\mathbb{N}}{k}$ under colex are precisely the subsets of $[\ell]$ of cardinality k. Thus, the result follows from Theorem 2.26. $\square$

Let us now show how the Kruskal–Katona Theorem (in the form of Corollary 2.45) implies the Erdős–Ko–Rado Theorem. Given a family $\mathcal{F}$ and $t \geq 1$, let $\partial^{(t)}\mathcal{F}$ be the collection obtained by applying the shadow operation to $\mathcal{F}$ exactly t times. That is, $\partial^{(1)}\mathcal{F} = \partial\mathcal{F}$ and, for $t \geq 2$,

$$\partial^{(t)}\mathcal{F} := \partial\left(\partial^{(t-1)}\mathcal{F}\right).$$

Another Proof of Theorem 1.5. Suppose that $\mathcal{F} \subseteq \binom{[n]}{k}$ is intersecting. Define

$$\mathcal{G} := \{[n] \setminus A : A \in \mathcal{F}\}.$$

If $A \in \mathcal{F}$ and $B \in \mathcal{G}$, then we cannot have $A \subseteq B$. This is because $\mathcal{F}$ is intersecting and the set B is of the form $[n] \setminus A'$ for some $A' \in \mathcal{F}$. Therefore,

$$\mathcal{F} \cap \partial^{(n-2k)}\mathcal{G} = \emptyset.$$

Suppose that $|\mathcal{F}| \geq \binom{n-1}{k-1}$. Then, since $|\mathcal{G}| = |\mathcal{F}|$, we also have $|\mathcal{G}| \geq \binom{n-1}{k-1} = \binom{n-1}{n-k}$. By Corollary 2.45 and the fact that $\mathcal{G} \subseteq \binom{[n]}{n-k}$, we have that

$$|\partial\mathcal{G}| \geq \binom{n-1}{n-k-1}$$

and, by induction, for all $1 \leq t \leq n-k$,

$$|\partial^{(t)}\mathcal{G}| \geq \binom{n-1}{n-k-t}.$$

In particular, $|\partial^{(n-2k)}\mathcal{G}| \geq \binom{n-1}{k}$. However, since this set is disjoint from $\mathcal{F}$ (we already proved this above), this implies that $|\mathcal{F}| \leq \binom{n}{k} - \binom{n-1}{k} = \binom{n-1}{k-1}$. This completes the proof. $\square$

2.4 The Littlewood–Offord Problem

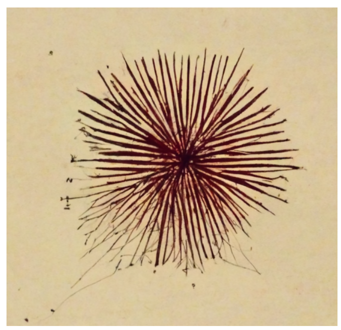


Figure 2.14: “Unit vectors.”

To close this chapter, we present an unexpected application of the ideas used in the proof of Sperner's Theorem. In a paper on the roots of random polynomials from 1943, Littlewood and Offord [177] asked the following.

Problem 2.46 (Littlewood Offord Problem [177])**.** *Given $z_1, \ldots, z_n \in \mathbb{C}$ and $A \subseteq [n]$, let*

$$z_A := \sum_{i \in A} z_i. \tag{2.47}$$

What is the largest cardinality of a collection $\mathcal{F} \subseteq 2^{[n]}$ for which there exists complex numbers $z_1, \ldots, z_n$ such that $|z_i| \geq 1$ for all $1 \leq i \leq n$ and

$$|z_A - z_B| < 1$$

for all $A, B \in \mathcal{F}$?

A couple of years later, Erdős [91] noticed a connection to Sperner's Theorem, at least in the special case that the numbers z_i are confined to the real numbers. First of all, if $z_1 = z_2 = \cdots =$

$z_n = 1$, then $\mathcal{F} \subseteq 2^{[n]}$ satisfies the property in Problem 2.46 if and only if all sets in $\mathcal{F}$ have the same size; thus, by taking $\mathcal{F}$ to be $\binom{[n]}{\lfloor n/2 \rfloor}$, we have a lower bound of $\binom{n}{\lfloor n/2 \rfloor}$.

Now, let us prove a matching upper bound (in the real case). Suppose that all of $z_1, \ldots, z_n$ are real numbers of absolute value at least one and let $\mathcal{F}$ be a collection satisfying the property in Problem 2.46. Note that $\mathcal{F}$ has that property if and only if there is a point $z \in \mathbb{R}$ such that

$$|z_A - z| < 1/2$$

for all $A \in \mathcal{F}$. If one of the z_i is negative, we replace z_i by $-z_i$, replace each element A of $\mathcal{F}$ with $A \triangle \{i\}$ (where $\triangle$ denotes the *symmetric difference*) and replace z by $z - z_i$. So, we can assume that all of the z_i are real numbers bounded below by 1. To finish the argument, we claim that $\mathcal{F}$ is an antichain. This is because, for $A \subsetneq B$, we have

$$1 \leq z_A - z_B = |z_A - z_B| \leq |z_A - z| + |z_B - z|$$

and so one of the terms on the right side must be at least $1/2$. Thus, $\mathcal{F}$ is an antichain and, by Sperner's Theorem, it has at most $\binom{n}{\lfloor n/2 \rfloor}$ elements.

This short and sweet argument settles the case of real numbers, but the original question of Littlewood and Offord for complex numbers remained unsolved until Kleitman [153] and Katona [143] showed that the answer was $\binom{n}{\lfloor n/2 \rfloor}$ in the mid-1960s. Later, Kleitman [155] gave a short and sweet proof that $\binom{n}{\lfloor n/2 \rfloor}$ remains the answer even when $z_1, \ldots, z_n$ are chosen from any normed space, which we present next. We assume that the definition of z_A in (2.47) is extended to the case where the points z_i are in any vector space (not just $\mathbb{C}$).

Theorem 2.48 (Kleitman [153])**.** *Let V be a normed vector space. If $z_1, \ldots, z_n$ are elements of V of norm at least one and $\mathcal{F} \subseteq 2^{[n]}$ such that*

$$\|z_A - z_B\| < 1$$

for all $A, B \in \mathcal{F}$, then $|\mathcal{F}| \leq \binom{n}{\lfloor n/2 \rfloor}$.

Proof. Rather than applying Sperner's Theorem, as is done in Erdős' proof for real numbers, let us instead mimic the proof of Sperner's Theorem given earlier in the notes, except with a more flexible type of structure taking the place of a symmetric chain decomposition.

First, let us modify the notion of a "chain" to fit our needs. Say that a family $\mathcal{D} \subseteq 2^{[n]}$ is *separated* if

$$\|z_A - z_B\| \geq 1$$

for all distinct $A, B \in \mathcal{D}$. This is the antithesis of a family satisfying the property described in the theorem, in the same way that a chain is the antithesis of an antichain.

Next, we loosen the notion of symmetric. Say that a partition $\mathcal{D}_1, \ldots, \mathcal{D}_s$ of $2^{[n]}$ is *appropriately sized* if, for $0 \leq i \leq \lfloor n/2 \rfloor$, the number of indices j such that $|\mathcal{D}_j| = n + 1 - 2i$ is precisely $\binom{n}{i} - \binom{n}{i-1}$; note that, here, we view $\binom{n}{-1}$ as being equal to zero. Our aim is to show that $2^{[n]}$ has an appropriately sized partition $\mathcal{D}_1, \ldots, \mathcal{D}_s$ such that each of the sets $\mathcal{D}_j$ is separated by induction on n.

The base case $n = 1$ is easy, and so consider $n \geq 2$. Let $\mathcal{D}_1, \ldots, \mathcal{D}_s$ be an appropriately sized partition of $2^{[n-1]}$ into separated families. Let $f : V \to \mathbb{R}$ be a linear map such that $f(z_n) = \|z_n\|$ and $f(x) \leq \|x\|$ for all $x \in V$. For example, if V is a real vector space, then we can let π be the linear projection of V onto $\mathrm{span}(\{z_n\})$, let ℓ map αz_n to $\alpha\|z_n\|$ for all $\alpha \in \mathbb{R}$, and then take f to be the composition of π and ℓ. The complex case isn't much harder. For each j with $1 \leq j \leq s$, let A_j be a set such that $A_j \in \mathcal{D}_j$ and, subject to this, the value of

$$f\left(z_{A_j}\right)$$

is maximized. Let

$$\mathcal{D}'_j := \mathcal{D}_j \cup \{A_j \cup \{n\}\}$$

and

$$\mathcal{D}''_j := \{B \cup \{n\} : B \in \mathcal{D}_j \text{ and } B \neq A_j\}.$$

By construction, the collections $\mathcal{D}'_1, \dots, \mathcal{D}'_s, \mathcal{D}''_1, \dots, \mathcal{D}''_s$ are disjoint and union to $2^{[n]}$. Now, how many of the families in $\mathcal{D}'_1, \dots, \mathcal{D}'_s, \mathcal{D}''_1, \dots, \mathcal{D}''_s$ have cardinality $n-2i+1$? We have that $\mathcal{D}'_j$ has this size if and only if $\mathcal{D}_j$ had size $n-2i$; there are $\binom{n-1}{i} - \binom{n-1}{i-1}$ such indices j. If $\mathcal{D}''_j$ has size $n-2i+1$, then the size of $\mathcal{D}_j$ was $n-2i+2$; the number of indices for which this is the case is $\binom{n-1}{i-1} - \binom{n-1}{i-2}$. Thus, by Pascal's Formula,[5] we have constructed an appropriately sized partition.

Now, it is easy to see that the collection $\mathcal{D}''_j$ is separated, as it is simply a translation of a subset of $\mathcal{D}_j$. As for $\mathcal{D}'_j$, all of its elements apart from $A_j \cup \{n\}$ were in $\mathcal{D}_j$. So, to check that it is separated, we let B be any set in $\mathcal{D}_j$ and observe that, by definition of f,

$$\left\|z_{A_j \cup \{n\}} - z_B\right\| \geq f\left(z_{A_j \cup \{n\}} - z_B\right)$$

$$= f\left(z_{A_j}\right) - f\left(z_B\right) + f(z_n) \geq f(z_n) = \|z_n\| \geq 1$$

where, here, we used the specific choice of the set A_j.

Let $\mathcal{D}^*_1, \dots, \mathcal{D}^*_t$ be the non-empty families among $\mathcal{D}'_1, \dots, \mathcal{D}'_s, \mathcal{D}''_1, \dots, \mathcal{D}''_s$. By assumption, $\mathcal{F}$ cannot intersect any of $\mathcal{D}^*_1, \dots, \mathcal{D}^*_t$ in more than one element. Thus,

$$|\mathcal{F}| = \sum_{j=1}^{t} |\mathcal{F} \cap \mathcal{D}^*_j| \leq t.$$

To conclude, we upper bound t as follows:

$$t \leq \sum_{i=0}^{\lfloor n/2 \rfloor} |\{j : |\mathcal{D}^*_j| = n-2i+1\}| = \sum_{i=0}^{\lfloor n/2 \rfloor} \binom{n}{i} - \binom{n}{i-1} = \binom{n}{\lfloor n/2 \rfloor}$$

since this sum is telescoping and $\binom{n}{-1} = 0$. $\square$

2.5 Exercises

2.1 List all of the antichains in $2^{[n]}$ for each $n \in \{1, 2, 3\}$.

2.2 In the proof of Lemma 2.13, show that $\mathcal{C}'_m$ and $\mathcal{C}''_m$ are symmetric chains in $2^{[n]}$ and that every subset of $[n]$ belongs to exactly one of the collections $\mathcal{C}'_1, \dots, \mathcal{C}'_N, \mathcal{C}''_1, \dots, \mathcal{C}''_N$.

[5] See the notes of Math 422 at UVic [68, p. 6].

2.3 In the proof of Lemma 2.13, it appears that the number of chains used in the partitioning of $2^{[n]}$ is exactly twice as many as were used in the partitioning of $2^{[n-1]}$. By induction, this suggests that $2^{[n]}$ can be covered by exactly 2^{n-1} symmetric chains. However, from the proof of Theorem 2.5, we know that the number of chains in any symmetric chain decomposition is exactly $\binom{n}{\lfloor n/2 \rfloor}$, which is certainly not equal to 2^{n-1} in general. Explain why the number of chains in the decomposition of $2^{[n]}$ found in the proof of Lemma 2.13 is sometimes less than twice the number of chains in the decomposition of $2^{[n-1]}$.

2.4 Let $\mathcal{C}_1, \dots, \mathcal{C}_N$ be a symmetric chain decomposition of $2^{[n]}$ where $N = \binom{n}{\lfloor n/2 \rfloor}$. For $0 \leq k \leq n/2$, how many chains in the decomposition have length exactly $n + 1 - 2k$?

2.5 (a) For $r \geq 2$, let $\mathcal{P}_r^n$ be the partially ordered sets defined in Exercise 1.27. For $0 \leq k \leq (r-1)n$, let
$$L_{r,k} := \left\{ (x_1, \dots, x_n) \in \mathcal{P}_r^n : \sum_{i=1}^n x_i = k \right\}$$
A *chain* in $\mathcal{P}_r^n$ is a set $C \subseteq \mathcal{P}_r^n$ such that $x \preceq y$ or $y \preceq x$ for all $x, y \in C$. A *symmetric chain* is a chain C such that there exists $0 \leq k \leq (r-1)n/2$ such that C contains an element of L_i for all $k \leq i \leq (r-1)n - k$. Prove that $\mathcal{P}_r^k$ can be partitioned into symmetric chains.

(b) Use the result of part (a) to determine the size of the largest antichain in $\mathcal{P}_r^n$, where an *antichain* in $\mathcal{P}_r^n$ is a set $A \subseteq \mathcal{P}_r^n$ such that $x \not\preceq y$ for any distinct $x, y \in A$.

2.6 (a) For $r \geq 3$ and $0 \leq k \leq (r-1)n$, let $\mathcal{P}_r^n$ and $L_{r,k}$ be as in Exercise 2.5. In this exercise, we focus on the case $r = 3$. Prove that, for every $1 \leq k \leq 2n$, there exists a function $f_k : L_{3,k} \times L_{3,k-1} \to [0, \infty)$ satisfying the following two properties:
$$\sum_{y \in L_{3,k-1}} f_k(x, y) = |L_{3,k-1}| \text{ for all } x \in L_{3,k},$$
$$\sum_{x \in L_k} f_k(x, y) = |L_{3,k}| \text{ for all } y \in L_{3,k-1}.$$

(b) A chain C in $\mathcal{P}_r^n$ is *maximal* if it contains an element of $L_{r,k}$ for all $0 \leq k \leq (r-1)n$. Prove that there exists a probability distribution on the set of maximal chains in $\mathcal{P}_3^n$ with the property that, if C is chosen randomly according to this distribution, then, for all $0 \leq k \leq 2n$ and $x \in L_{3,k}$, it holds that $\mathbb{P}(x \in C) = \frac{1}{|L_{3,k}|}$.

(c) State and prove a version of the LYM Inequality for $\mathcal{P}_3^n$.

(d) State and prove a version of the Local LYM Inequality for $\mathcal{P}_3^n$.

2.7 Let $0 \leq r \leq n/2$ and let $\mathcal{F} \subseteq 2^{[n]}$ be an intersecting antichain such that every set in $\mathcal{F}$ has cardinality at most r. Prove that $|\mathcal{F}| \leq \binom{n-1}{r-1}$.

2.8 a) Let A and B be disjoint subsets of $[n]$. Consider a random permutation of $[n]$. What is the probability that all of the elements of A come before all of the elements of B in the permutation?

b) Let $(A_1, B_1), \ldots, (A_m, B_m)$ be a sequence of pairs of finite sets such that $A_i \cap B_i = \emptyset$ and $A_i \cap B_j \neq \emptyset$ whenever $i \neq j$. Prove that

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

c) Use the statement in the previous part of the question to prove the LYM Inequality.

2.9 Let k be a fixed positive integer and let n be an integer which we view as tending to infinity. Let $\mathcal{F} \subseteq 2^{[n]}$ be an antichain containing at least one set of cardinality at most k, at least one of cardinality at least $n-k$ and none of cardinality strictly between k and $n-k$. Prove that $|\mathcal{F}| = O(n^{k-1})$.

2.10 Say that $\mathcal{F} \subseteq \binom{[n]}{k}$ is *intersection-free* if there does not exist distinct $A, B, C \in \mathcal{F}$ such that $A \cap B \subseteq C$. Prove that if $\mathcal{F} \subseteq \binom{[n]}{k}$ is intersection-free, then $|\mathcal{F}| \leq 1 + \binom{k}{\lfloor k/2 \rfloor}$.

2.11 Let $\mathcal{F} \subseteq 2^{[n]}$ such that $|\mathcal{F}| = m$ and, for all distinct $i, j \in [n]$, there exists $A \in \mathcal{F}$ containing i but not j. Prove that $n \leq \binom{m}{\lfloor m/2 \rfloor}$.

2.12 For each set $A \subseteq [n]$, define $\mu(A) := (-1)^{|A|}$. Say that a collection $\mathcal{F} \subseteq 2^{[n]}$ is *convex* if for any sets $A, B, C \subseteq [n]$ such that $A \subseteq C \subseteq B$, if $A, B \in \mathcal{F}$, then $C \in \mathcal{F}$. Prove that, if $\mathcal{F}$ is convex, then

$$\left| \sum_{A \in \mathcal{F}} \mu(A) \right| \leq \binom{n}{\lfloor n/2 \rfloor}.$$

2.13 Let $f(n)$ be the number of antichains in $2^{[n]}$.

(a) Prove that $f(n) \geq 2^{\binom{n}{\lfloor n/2 \rfloor}}$.

(b) Prove that, for every $\varepsilon > 0$, there exists $N(\varepsilon) \in \mathbb{N}$ such that if $n \geq N(\varepsilon)$, then $f(n) \leq 2^{\varepsilon 2^n}$.

Hint 1: You may want to use $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.

Hint 2: Look up Stirling's Approximation and apply it.[6]

2.14 Let $k \leq n/2$ and suppose that $\mathcal{A}$ is an antichain in $2^{[n]}$ in which every element of $\mathcal{A}$ has cardinality at most k. Prove that $|\mathcal{A}| \leq \binom{n}{k}$.

2.15 For $n \geq 1$, a *rising antichain* is a collection $\mathcal{A} \subseteq 2^{[n]}$ such that $\mathcal{A}$ is an antichain and no two elements of $\mathcal{A}$ have the same cardinality.

(a) Show that, for every $n \geq 1$, every rising antichain in $2^{[n]}$ has cardinality at most $n+1$ (Easy).

(b) Determine, for every $n \geq 1$, the maximum possible size of a rising antichain in $2^{[n]}$.

[6] One place in which you can find Stirling's Approximation is in the course notes to Math 422 at UVic [68, Theorem 1.7].

2.16 (a) Prove that, if $\mathcal{F} \subseteq 2^{[n]}$ does not contain a chain of cardinality k, then there exists a partition of $\mathcal{F}$ into at most $k-1$ antichains. (This is known as *Mirsky's Theorem*)

(b) Using part (a) and the LYM Inequality, prove Corollary 2.16.

2.17 Use the Local LYM Inequality to prove that, if $\mathcal{F} \subseteq 2^{[n]}$ is a downset, then the average cardinality of an element in $\mathcal{F}$ is at most $n/2$.

2.18 Use the Local LYM Inequality to prove the LYM Inequality.

2.19 Let $w : \{0, 1, \dots, n\} \to \mathbb{R}$ be any *weight function* and, for $\mathcal{A} \subseteq 2^{[n]}$, let

$$w(\mathcal{A}) := \sum_{A \in \mathcal{A}} w(|A|).$$

Prove that, for any antichain $\mathcal{A} \subseteq 2^{[n]}$,

$$w(\mathcal{A}) \leq \max_{0 \leq k \leq n} \binom{n}{k} w(k).$$

2.20 Suppose that $\mathcal{F} \subseteq 2^{[n]}$ has the property that, for any $A \in \mathcal{F}$, there does not exist a pair of sets $B, C \in \mathcal{F}$ such that $B \subsetneq A \subsetneq C$. Prove that $|\mathcal{F}|$ is at most $2\binom{n}{\lfloor n/2 \rfloor}$. For which values of n is this tight?

2.21 Say that an antichain $\mathcal{A} \subseteq 2^{[n]}$ is *complementary* if for every $A \in \mathcal{A}$, the set $[n] \setminus A$ is also in $\mathcal{A}$. Show that the maximum cardinality of a complementary antichain is $2\binom{n-1}{\lceil n/2 \rceil}$ for every $n \geq 1$.

2.22 For each k such that $0 \leq k \leq \binom{n}{\lfloor (n+2)/2 \rfloor}$, give an example of a collection $\mathcal{A} \subseteq 2^{[n]}$ with $|\mathcal{A}| = \binom{n}{\lfloor n/2 \rfloor} + k$ such that the number of pairs $A, B \in \mathcal{A}$ satisfying $A \subsetneq B$ is equal to $k \left\lfloor \frac{n+2}{2} \right\rfloor$.

2.23 Say that a collection $\mathcal{F} \subseteq 2^{[n]}$ is *k-chain saturated* if it does not contain a k-chain but, for any $A \in 2^{[n]} \setminus \mathcal{F}$, the collection $\mathcal{F} \cup \{A\}$ does contain a k-chain. Prove that, if $n \geq k-2$, then there exists a k-chain saturated family $\mathcal{F} \subseteq 2^{[n]}$ such that $|\mathcal{F}| = 2^{k-2}$.

2.24 Write out $\binom{[5]}{3}$ ordered by colex.

2.25 What are the 99th, 100th and 101st elements in the colex order on $\binom{\mathbb{N}}{4}$? What about lex order?

2.26 Let $\mathcal{A} \subseteq \binom{[9]}{3}$ with $|\mathcal{A}| = 28$. How small can $|\partial \mathcal{A}|$ be?

2.27 The *upper shadow* of a family $\mathcal{F} \subseteq \binom{[n]}{k}$ is

$$\partial^+ \mathcal{F} := \{A \subseteq [n] : |A| = k+1 \text{ and } A \supseteq F \text{ for some } F \in \mathcal{F}\}.$$

State and prove a version of the Kruskal–Katona Theorem for the upper shadow.

2.28 Give an example of a finite set system $\mathcal{F} \subseteq \binom{\mathbb{N}}{2}$ which is not isomorphic to the collection $\mathcal{I}$ of the first $|\mathcal{F}|$ sets of $\binom{\mathbb{N}}{2}$ under colex order such that $|\partial \mathcal{F}| = |\partial \mathcal{I}|$.

2.6 Challenge Problems

2.1* Prove that, for any $\mathcal{A} \subseteq 2^{[n]}$, the number of pairs $A, B \in \mathcal{A}$ such that $A \subsetneq B$ is at least

$$\left(|\mathcal{A}| - \binom{n}{\lfloor n/2 \rfloor}\right) \left\lfloor \frac{n+2}{2} \right\rfloor .$$

Hint: Expand upon the "random chain" argument from the proof of the LYM Inequality.

2.2* Prove that, for any $k \geq 0$ and $\mathcal{A} \subseteq 2^{[n]}$ with $|\mathcal{A}| = \binom{n}{\lfloor n/2 \rfloor} + \binom{n}{\lfloor (n+2)/2 \rfloor} + k$, the number of pairs $A, B \in \mathcal{A}$ such that $A \subsetneq B$ is $\Omega(k \cdot n^2)$.

2.3* Prove that there exists a constant $C > 0$ such that the number of antichains in $2^{[n]}$ is at most $C^{\binom{n}{\lfloor n/2 \rfloor}}$ for every n.

2.4* For $0 \leq k \leq n-1$, let $\mathcal{A} \subseteq \binom{[n]}{k}$ and $\mathcal{B} \subseteq \binom{[n]}{k+1}$ be initial segments of colex order with $|\mathcal{A}| = |\mathcal{B}|$. Does it always hold that $|\partial \mathcal{A}| \leq |\partial \mathcal{B}|$?

Chapter 3

Classical Extremal Graph Theory

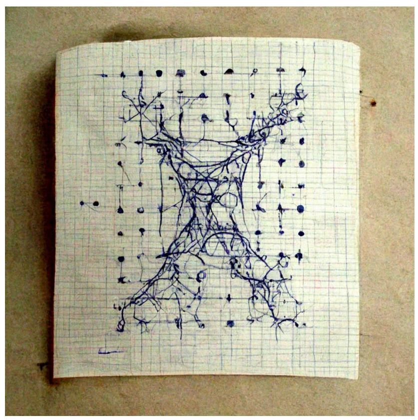


Figure 3.1: "Graph Theory."

Recall that a *graph* is a pair $G = (V, E)$ such that V is a set of *vertices* and E is a collection of unordered pairs of vertices called *edges*. Given a graph G, we often denote its vertex set by $V(G)$ and edge set by $E(G)$. We will always consider *simple graphs*; i.e. graphs without loops or multiple edges.

Given a graph G and $v \in V(G)$, say that $u \in V(G)$ is a *neighbour* of v or is *adjacent* to v if $uv \in E(G)$. Let $N(v)$ be the set of neighbours of v and $d(v) := |N(v)|$ be the *degree* of v. We sometimes write $N_G(v)$ or $d_G(v)$ if the graph is not clear from context. Let $\delta(G) := \min\{d(v) : v \in V(G)\}$ and $\Delta(G) := \max\{d(v) : v \in V(G)\}$. Given a set $S \subseteq V(G)$, the subgraph of G *induced* by S is the graph $G[S]$ with vertex set S and edge set $\{uv \in E(G) : u, v \in S\}$. Given sets A and B, let $e(A)$ be the number of edges of G contained in A and $e(A, B)$ be the number of of edges with one endpoint in A and one endpoint in B; again, we may write e_G when doing so would help to avoid confusion. A subgraph of a graph G is a graph whose vertex and edge sets are subsets of those of G. Two graphs G_1 and G_2 are *isomorphic* if there is a bijection $f : V(G_1) \to V(G_2)$ such that $uv \in E(G_1)$ if and only if $f(u)f(v) \in E(G_2)$.

Say that a *copy* of a graph H in a graph G is a subgraph of G that is isomorphic to H. The *number of copies* of H in G is therefore the number of such subgraphs. For example, the number of copies of K_2, the complete graph on two vertices, in any graph G is exactly $|E(G)|$. Say that a graph G is *H-free* if it does not contain a copy of H.

A large part of extremal graph theory is focused on understanding the "interaction" between the number of copies of "small" graphs $H_1, H_2, \ldots, H_k$ within a "large" graph G. For instance, a typical question in this area is

> Given a graph H and an integer $n \in \mathbb{N}$, what is the maximum possible number of edges in an H-free graph with n vertices?

We denote this number by $\mathrm{ex}(n, H)$ and call it the *extremal number* of H (or *nth extremal number*, if such precision is necessary). The focus of this chapter and the one which follows it is on determining, or estimating, the value of $\mathrm{ex}(n, H)$ for different graphs H. Some extremal graph theory (e.g. Turán's Theorem) is usually covered in Math 423 at UVic. However, we will go into greater depth and take a somewhat different approach.

3.1 The Extremal Number of a Triangle

The *complete graph* on r vertices is the graph K_r with r vertices in which every pair of distinct vertices are adjacent. A complete graph is often called a *clique*. In this section and the one that follows it, we will consider the classical problem of determining $\mathrm{ex}(n, K_r)$ for $r \geq 3$. As a warm-up, let us start by considering the special case $r = 3$ (which is known as *Mantel's Theorem*). A K_3-free graph is often referred to as a *triangle-free* graph.

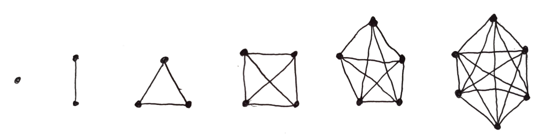


Figure 3.2: The complete graphs K_1, K_2, K_3, K_4, K_5 and K_6.

Example 3.1. A graph G is *bipartite* if $V(G)$ can be partitioned into two sets A and B such that every edge of G has an endpoint in A and an endpoint in B. We say that G is a *complete bipartite graph* if it contains every edge from A to B; we write $G = K_{s,t}$ if it is complete bipartite with $|A| = s$ and $|B| = t$.

It is clear that a bipartite graph cannot contain a copy of K_3; indeed, it can only contain copies of other bipartite graphs, which K_3 is not. This makes a complete bipartite graph a natural choice for trying to maximize the number of edges in a triangle-free graph. The graph $K_{s,t}$ clearly contains st edges. Thus, among all choices of s and t with $s + t = n$, the number of edges in $K_{s,t}$ is maximized when s and t are as similar as possible. Therefore,

$$\mathrm{ex}(n, K_3) \geq \left\lfloor \frac{n}{2} \right\rfloor \left\lceil \frac{n}{2} \right\rceil = \left\lfloor \frac{n^2}{4} \right\rfloor$$

for all n.

As it turns out, Example 3.1 is tight.

Figure 3.3: The complete bipartite graphs $K_{3,3}$ and $K_{2,4}$.

Theorem 3.2 (Mantel's Theorem [193]). *If G is a triangle-free graph, then*

$$|E(G)| \leq \frac{|V(G)|^2}{4}.$$

Consequently,

$$\operatorname{ex}(n, K_3) = \left\lfloor \frac{n^2}{4} \right\rfloor$$

for all n.

We give a proof of Mantel's Theorem based on the following lemma and a simple convexity argument (see Appendix C). It is important to get a good grasp of what is going on in this proof, as a similar approach will be used in several of the proofs in the rest of this chapter and the one which follows it.

Lemma 3.3. *For any graph G,*

$$\sum_{v \in V(G)} d(v)^2 = \sum_{uv \in E(G)} (d(u) + d(v)).$$

Proof. We give a "combinatorial proof."[1] That is, our aim is to show that the quantities on the left and right sides of the equality that we are proving "count" the same set of "objects;" if this is true, then they must be equal.

The objects that we consider are triples $(x, y_1, y_2) \in V(G)^3$ such that xy_1 and xy_2 are both in $E(G)$. Perhaps the simplest way to count these objects is to

1. select a vertex $x \in V(G)$,
2. select a neighbour y_1 of x and
3. select another neighbour y_2 of x.

Note that we allow y_1 and y_2 to be the same vertex. It is not hard to see that performing these three steps precisely counts the triples (x, y_1, y_2) that we are aiming to count. Thus, the number of such triples is

$$\sum_{x \in V(G)} d(x)^2$$

which is the same as the left side of the equality in the lemma. Now, suppose that we do the following instead:

[1] Anyone who is taking, or has taken, Math 422 at UVic will be familiar with the concept of a combinatorial proof; see [68, Section 1.1].

1. select an edge $e \in E(G)$,

2. select one of the endpoints of e and call it x,

3. let the other endpoint of e be called y_1,

4. select another neighbour y_2 of x.

Any triple (x, y_1, y_2) of the type that we are trying to count will be chosen if and only if the edge selected in the first step is xy_1, the vertex selected in the second step is x and the vertex chosen in the fourth step is y_2. Thus, this procedure counts every such triple exactly once. Thus, the number of such triples is also equal to

$$\sum_{uv \in E(G)} (d(u) + d(v)).$$

□

Proof of Mantel's Theorem. Let G be a triangle-free graph. Since G is triangle-free, we have that for any edge $uv \in E(G)$, each vertex $w \in V(G)$ can be adjacent to one of u or v but not both; note, this holds even in the case that $w = u$ or $w = v$ as no vertex is adjacent to itself. Therefore,

$$d(u) + d(v) \leq |V(G)|$$

for any $uv \in E(G)$. By Lemma 3.3 and the above inequality,

$$\sum_{v \in V(G)} d(v)^2 = \sum_{uv \in E(G)} (d(u) + d(v)) \leq |E(G)| \cdot |V(G)|.$$

By the standard "handshaking lemma," we have

$$\sum_{v \in V(G)} d(v) = 2|E(G)|.$$

Putting this all together and applying Corollary C.7, we obtain

$$4|E(G)|^2 = \left(\sum_{v \in V(G)} d(v) \right)^2 \leq |V(G)| \cdot \sum_{v \in V(G)} d(v)^2 \leq |V(G)|^2 \cdot |E(G)|.$$

The result follows by dividing both sides by $4|E(G)|$. □

3.2 The Extremal Numbers of General Cliques

Next, let us consider the problem of determining $\text{ex}(n, K_r)$ for general r. The following is the natural generalization of the tight construction for Mantel's Theorem.

Example 3.4. For $k \geq 1$, a *complete k-partite graph* is a graph such that the vertex set can be partitioned into k sets $A_1, \ldots, A_k$, called *parts*, such that there is an edge from every $u \in A_i$ to every $v \in A_j$ for $i \neq j$, but there are no edges between vertices of A_i. We denote this graph by $K_{a_1, a_2, \ldots, a_k}$ where $a_i = |A_i|$ for $1 \leq i \leq k$.

It is clear that no $(r-1)$-partite graph contains a copy of K_r. Assuming the number of edges is n, the number of edges in such a graph is maximized when every part is of size $\lfloor n/(r-1) \rfloor$ or $\lceil n/(r-1) \rceil$; see Exercise 3.12. Such graphs are often referred to as *Turán graphs.*

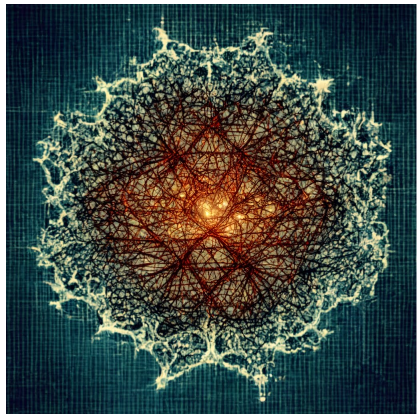


Figure 3.4: “A dense graph.”

Again, we will show that this construction is optimal. Note that the following theorem implies Mantel's Theorem, but the proof that we will give, when specialized to the case $r = 3$, will be different than the one given above. Thus, it gives an alternative proof.

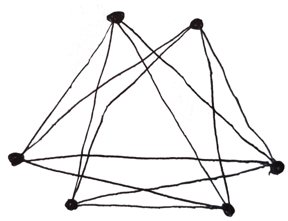


Figure 3.5: The graph $K_{2,2,2}$.

Theorem 3.5 (Turán's Theorem [262])**.** *Let $n \geq r \geq 3$ be integers and let*

$$t = n - (r-1)\left\lfloor \frac{n}{r-1} \right\rfloor .$$

Then $\text{ex}(n, K_r)$ is equal to the number of edges in a complete $(r-1)$-partite graph with $r-1-t$ parts of size $\lfloor n/(r-1) \rfloor$ and t parts of size $\lceil n/(r-1) \rceil$.

Proof. What we will prove is that the maximum number of edges in a K_r-free graph is attained by a complete $(r-1)$-partite graph; the fact that all of the parts have nearly the same size is a simple optimization (see Exercise (a)).

Given a graph G and a vertex $v \in V(G)$, let $G * v$ be the graph obtained from G by adding a new vertex, say v', and adding an edge from v' to every neighbour of v. We call this operation *cloning* the vertex v. The proof relies on the following claim.

Claim 3.6. *Let G be a graph and $v \in V(G)$. If G is K_r-free and, then so is $G * v$.*

Proof. Exercise 3.11. □

Let G be a K_r-free graph on n vertices with the maximum possible number of edges. Let u and v be a pair of non-adjacent vertices such that $d(v) \geq d(u)$. Consider the graph

$$G' = (G - u) * v.$$

If you jettison a vertex from a K_r-free graph, then it remains K_r-free; thus, $G - u$ is K_r-free. So, by Claim 3.6, G' is K_r-free, too. Since u and v are not adjacent, the number of edges in G' is $|E(G)| - d(u) + d(v)$. Thus, by our choice of G, it must hold that $d(v) = d(u)$; so, any two non-adjacent vertices have the same degree.

Now, suppose that there is a triple of vertices u, v, w such that $uv, vw \notin E(G)$ and $uw \in E(G)$. By the result of the previous paragraph, we have that $d(u) = d(v) = d(w)$. Let

$$G'' = ((G - u - w) * v) * v.$$

This is a K_r-free graph with the number of edges equal to

$$|E(G)| - d(u) - (d(w) - 1) + 2d(v) = |E(G)| + 1$$

(since, after deleting u, the degree of w decreases by one, but the degree of v is not affected by deleting u or w). This contradicts our choice of G; thus, no triple of this type can exist.

From the previous paragraph, we see that no two vertices in the same component of the complement $\overline{G}$ of G can be at distance greater than one from one another (otherwise, the closest such vertices are at distance two, which gives us a triple as in the previous paragraph). Therefore, G is a complete multipartite graph. □

Remark 3.7. Note that the proof of Turán's Theorem given above also implies that the extremal graph for Turán's Theorem is unique.

3.3 Turán Density of General Graphs

Figure 3.6: "UVic math student during midterm season."

So far, we have only considered extremal numbers of complete graphs, which begs the question of how extremal numbers behave for graphs in general. In this section, we will see that, for most graphs, it is possible to pin down the extremal number, up to some small order fluctuations.

An *independent set* in a graph G is a set $S \subseteq V(G)$ such that there does not exist $u, v \in S$ such that $uv \in E(G)$. One of the keys to computing the extremal numbers of complete graphs in the previous section was that a complete graph on r vertices cannot be decomposed into fewer than r independent sets. This property turns out to be important for estimating extremal numbers in gener. Given a graph H, the *chromatic number* of H, denoted $\chi(H)$, is the minimum k such that $V(H)$ can be partitioned into k independent sets; in other words, it is the smallest k such that the vertices of H can be coloured with K colours such that the endpoints of each edge are coloured differently. See Figure 3.3 for an example.

Example 3.8. For any graph H, a complete multipartite graph G with $\chi(H) - 1$ parts cannot contain a copy of H. Therefore,

$$\mathrm{ex}(n, H) \geq \mathrm{ex}\left(n, K_{\chi(H)}\right)$$

for all n. Asymptotically speaking, this tells us that

$$\mathrm{ex}(n, H) \geq \left(\frac{\chi(H) - 2}{\chi(H) - 1} + o(1)\right) \frac{n^2}{2}.$$

Definition 3.9. For $k \geq 1$, the *k-vertex path* is the graph P_k with k vertices $v_1, \ldots, v_k$ such that $v_i v_{i+1} \in E(P_k)$ for $1 \leq i \leq k - 1$.

Definition 3.10. For $k \geq 3$, the *k-cycle* is the graph C_k with k vertices $v_1, \ldots, v_k$ such that $E(P_k) \subseteq E(C_k)$ and $v_1 v_k \in E(C_k)$. That is, the vertices are joined in a "cyclic" fashion.

Figure 3.7: The graphs P_1, P_2, P_3 and P_4.

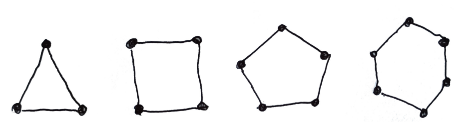


Figure 3.8: The graphs C_3, C_4, C_5 and C_6.

Example 3.11. Consider C_5. It is easy to see that $\chi(C_5) = 3$. For $n = 3$, it is clear that $\mathrm{ex}(3, C_5) = 3 > 2 = \mathrm{ex}(3, K_3)$; see Exercise 3.15. Thus, the bound in Example 3.8 may not be tight, at least for small values of n.

Example 3.12. Let H be a graph with the property that $\chi(H-e)=\chi(H)$ for every edge e of H. For a concrete example, one could take H to be a complete k-partite graph in which every part has at least two elements. Let G be a graph obtained from a complete $(\chi(H)-1)$-partite graph by adding a single edge within one of the parts. Then the vertices of the graph G can be coloured with $\chi(H)-1$ colours in such a way that only one edge is monochromatic; thus, G cannot contain any copy of H. Therefore, for any n, we have that

$$\mathrm{ex}(n,H)\geq \mathrm{ex}\left(n,K_{\chi(H)}\right)+1.$$

So, for any graph H of this type, the lower bound in Example 3.11 is never tight.

Examples 3.11 and 3.12 suggest that it may be tricky to determine the exact extremal number of H for all values of n. For each value of n, the exact structure of the extremal graphs G on n vertices may depend on the idiosyncrasies of the graph H that we are forbidding. However, what we will show is that, if all that we want to determine is the "rough" asymptotics of $\mathrm{ex}(n,H)$ (relative to n^2), then the fine details of the extremal constructions do not make a significant difference.

Definition 3.13. For each graph H, define the *Turán density* of H to be

$$\pi(H):=\lim_{n\to\infty}\frac{\mathrm{ex}(n,H)}{\binom{n}{2}}.$$

Of course, if we are defining $\pi(H)$ in terms of a limit, then the definition only makes sense if that limit exists.

Proposition 3.14. *For any graph H, $\pi(H)$ exists.*

Proof. Exercise 3.16. □

Remarkably, Erdős and Stone [89] exactly determined the Turán density of every graph.

Theorem 3.15 (Erdős–Stone Theorem [89])**.** *If H is any graph with at least one edge, then*

$$\pi(H)=\frac{\chi(H)-2}{\chi(H)-1}.$$

We divide the proof of this theorem into two lemmas. The first is a simple statement for converting a graph with large average degree into a graph with large minimum degree without deleting too many vertices.

Lemma 3.16. *Let G be a graph on $n\geq 1$ vertices and let $\alpha:=|E(G)|/\binom{n}{2}$. Then, for any $0<\gamma<\alpha$, there is a subgraph G' of G such that*

$$\delta(G')\geq(\alpha-\gamma)\left(|V(G')|-1\right)$$

and

$$|V(G')|\geq n\cdot\sqrt{\frac{\gamma}{1-\alpha+\gamma}}.$$

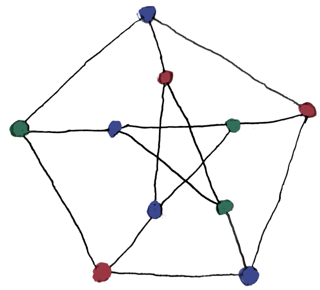


Figure 3.9: The Petersen graph P has chromatic number three. By the Erdős–Stone Theorem, $\mathrm{ex}(n,P) = (1+o(1))\frac{n^2}{4}$.

Proof. Initialize $G_0 := G$. For $i \geq 1$, let v_i be a vertex of $V(G_{i-1})$ of minimum degree in G_i and define G_i to be $G_{i-1} - v_i$. We terminate this process at the first index i such that

$$\delta(G_i) \geq (\alpha - \gamma)\left(|V(G_i)| - 1\right).$$

Note that this condition is eventually satisfied since the graph G_{n-1} has only one vertex, and so both sides of the above inequality are zero. Therefore, the termination condition is well-defined.

Now, by construction, for any $i \geq 1$ for which G_i is defined, it holds that

$$|E(G_{i-1})| = |E(G_i)| + \delta(G_{i-1}).$$

Thus,

$$|E(G)| = |E(G_i)| + \sum_{j=0}^{i-1} \delta(G_j).$$

If the process does not terminate before constructing G_i, then we must have $\delta(G_j) < (\alpha - \gamma)(|V(G_j)| - 1)$ for all $0 \leq j \leq i-1$. Therefore,

$$\begin{aligned} |E(G)| &< |E(G_i)| + \sum_{j=0}^{i-1}(\alpha-\gamma)(|V(G_j)|-1) \\ &\leq \binom{n-i}{2} + (\alpha-\gamma)\sum_{j=0}^{i-1}(n-1-j) = \binom{n-i}{2} + (\alpha-\gamma)\frac{i(2n-i-1)}{2} \\ &= \binom{n-i}{2} + (\alpha-\gamma)\left(\binom{n}{2} - \binom{n-i}{2}\right). \end{aligned}$$

Combining this with the hypothesis on G and doing a bit of re-arranging, we get that

$$\gamma\binom{n}{2} \leq (1-\alpha+\gamma)\binom{n-i}{2}.$$

Therefore, for any i such that G_i is defined, we have

$$\gamma(n-1)^2 < (1-\alpha+\gamma)(n-i)^2$$

$$\implies i < n - \sqrt{\frac{\gamma}{1-\alpha+\gamma}}(n-1) \leq \left(1 - \sqrt{\frac{\gamma}{1-\alpha+\gamma}}\right) n + 1.$$

Thus, when the process terminates, the number of vertices remaining is at least $n \cdot \sqrt{\frac{\gamma}{1-\alpha+\gamma}}$. This completes the proof. $\square$

The next lemma says that any graph with sufficiently large minimum degree contains a complete r-partite graph in which every part has cardinality t.

Lemma 3.17. *Let $r \geq 2$ and $t \geq 1$ and $\varepsilon > 0$. There exists $n_0(r, t, \varepsilon)$ such that, if G is a graph such that*

$$\delta(G) \geq \left(\frac{r-2}{r-1} + \varepsilon\right)(|V(G)| - 1)$$

and

$$|V(G)| \geq n_0(r, t, \varepsilon),$$

then G contains a complete r-partite graph with parts of cardinality t.

Proof. Define $t_r := t$ and, for $1 \leq s \leq r-1$, let $t_s := \lceil (2/\varepsilon) t_{s+1} \rceil$. What we will prove is that, for all $1 \leq s \leq r$, the graph G contains a complete s-partite graph in which each part has cardinality at least t_s. The result will then follow by setting $s = r$. We proceed by induction on s where, in the base case $s = 1$, we simply select any set of at least $t_1 \approx (2/\varepsilon)^{r-1} t$ vertices in G.

Now, for $s \geq 2$, let $B_1, \dots, B_{s-1}$ be the parts of a complete $(s-1)$-partite graph in G with parts of size exactly t_{s-1}. Define $B := \bigcup_{i=1}^{s-1} B_i$ and $U := V(G) \setminus B$. Define

$$W := \{u \in U : |N(u) \cap B_i| \geq t_s \text{ for all } 1 \leq i \leq s-1\}.$$

We want to prove that the set W is rather large. The way that we will do this is by "double counting" the non-edges from U to B. First of all, each vertex of $U \setminus W$ has more than

$$t_{s-1} - t_s \geq \left(1 - \frac{\varepsilon}{2}\right) t_{s-1}$$

non-neighbours in B. Thus, the number of non-edges between U and B is greater than

$$|U \setminus W| \left(1 - \frac{\varepsilon}{2}\right) t_{s-1} = (|V(G)| - |B| - |W|) \left(1 - \frac{\varepsilon}{2}\right) t_{s-1}.$$

On the other hand, by the minimum degree condition, each vertex in B is only incident to at most $\left(\frac{1}{r-1} - \varepsilon\right)(|V(G)| - 1)$ non-edges. Thus, the number of non-edges between U and B is at most

$$|B| \left(\frac{1}{r-1} - \varepsilon\right)(|V(G)| - 1) < t_{s-1}\left(1 - \varepsilon(r-1)\right)|V(G)|.$$

Putting this together, we get

$$(|V(G)| - |B| - |W|) \left(1 - \frac{\varepsilon}{2}\right) t_{s-1} < t_{s-1}\left(1 - \varepsilon(r-1)\right)|V(G)|$$

which implies that

$$|W| \left(1 - \frac{\varepsilon}{2}\right) > \varepsilon \cdot (r - (3/2)) \cdot |V(G)| - (r-1) t_{s-1}.$$

By letting $|V(G)|$ be sufficiently large, we get that

$$|W| > \binom{t_{s-1}}{t_s}^{s-1}(t_s - 1).$$

The number of ways to choose a set of size t_s from each of $B_1, \ldots, B_{s-1}$ is precisely $\binom{t_{s-1}}{t_s}^{s-1}$. Thus, by the Pigeonhole Principle,[2] there must be a set A of t_s vertices of W and subsets B_i' of B_i for $1 \leq i \leq s-1$ such that every vertex in A is adjacent to every vertex in $\bigcup_{i=1}^{s-1} B_i'$. This completes the proof. □

We now combine these two lemmas to prove the Erdős–Stone Theorem.

Proof of Theorem 3.15. Let H be any graph containing an edge, let $r = \chi(H)$ and let $t = |V(H)|$. Let $\varepsilon > 0$ and let n be large with respect to r, t and ε. Let G be a graph with n vertices such that

$$|E(G)| \geq \left(\frac{r-2}{r-1} + \varepsilon\right)\binom{n}{2}.$$

Using Lemma 3.16, we can find a subgraph G' of G such that

$$\delta(G') \geq \left(\frac{r-2}{r-1} + \frac{\varepsilon}{2}\right)(|V(G')| - 1)$$

and with the number of vertices in G' bounded below by $n\sqrt{\frac{\varepsilon}{2-\varepsilon}}$. Now, applying Lemma 3.17 to G', we see that, if n is large enough, then it contains a complete r-partite graph in which every part has cardinality t. This clearly contains H, and so we are done. □

3.4 Bounding Extremal Numbers of Bipartite Graphs

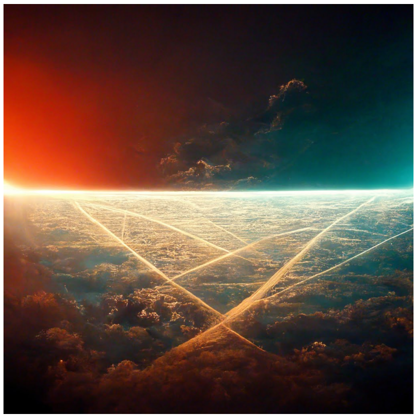


Figure 3.10: "Projective plane."

[2] See the course notes for Math 422 at UVic [68, Chapter 5].

One crucial thing to notice about the Erdős–Stone Theorem is that it tells us very little about the precise asymptotics of $\mathrm{ex}(n,H)$ when H is bipartite (i.e. when $\chi(H)=2$). That is, all that it tells us is that $\mathrm{ex}(n,H)$ grows slower than n^2, but it does not tell us what the actual growth rate is. The well-known Kővari–Sós–Turán Theorem tells us that it is at least a polynomial factor away from n^2.

Theorem 3.18 (Kővari–Sós–Turán Theorem [148])**.** *For any natural numbers s and t such that $s\leq t$,*

$$\mathrm{ex}(n,K_{s,t})\leq (1+o(1))\frac{1}{2}(t-1)^{\frac{1}{s}}n^{2-\frac{1}{s}}.$$

Proof. The proof has a similar flavour to the proof of Mantel's Theorem given earlier. Let G be a graph which does not contain a copy of $K_{s,t}$. We count the pairs (v,S) such that $v\in V(G)$ and S is a set of cardinality s contained in the neighbourhood of v. On one hand, this is clearly equal to

$$\sum_{v\in V(G)}\binom{d(v)}{s}.$$

On the other hand, for each set S of cardinality s there are at most $t-1$ such vertices v for, if not, then G would contain a $K_{s,t}$. Therefore,

$$\sum_{v\in V(G)}\binom{d(v)}{s}\leq (t-1)\binom{n}{s}.$$

Now, given the constraint that $\sum_{v\in V(G)}d(v)=2|E(G)|$, the sum $\sum_{v\in V(G)}\binom{d(v)}{s}$ is minimized when the degrees are as similar as possible; this is because, by Pascal's Formula,

$$\binom{a+1}{s}+\binom{b-1}{s}-\binom{a}{s}-\binom{b}{s}=\binom{a}{s-1}-\binom{b-1}{s-1}$$

which is non-negative whenever $a\geq b-1$. Thus,

$$n\cdot\frac{(2|E(G)|/n-s)^s}{s!}\leq\sum_{v\in V(G)}\binom{d(v)}{s}\leq (t-1)\binom{n}{s}\leq (t-1)\frac{n^s}{s!}.$$

The result now follows from a bit of algebra. □

Since any bipartite graph H is contained in a complete bipartite graph, we get the following.

Corollary 3.19. *For any bipartite graph H,*

$$\mathrm{ex}(n,H)=O\left(n^{2-\frac{2}{|V(H)|}}\right)$$

where the constant factor depends on H.

Consider the proof of Theorem 3.18 specialized to the case that $s=t=2$. Note that $K_{2,2}$ is the same as the 4-cycle C_4. If we fast forward to the middle of the proof, we have that, in any C_4-free graph G on n vertices,

$$\binom{n}{2}\geq\sum_{v\in V(G)}\binom{d(v)}{2}=\sum_{v\in V(G)}\frac{d(v)^2}{2}-\sum_{v\in V(G)}\frac{d(v)}{2}.$$

Recall that, by handshaking, $\sum_{v \in V(G)} d(v) = 2|E(G)|$. So, applying Corollary C.7, we have

$$\binom{n}{2} \geq \frac{(2|E(G)|)^2}{2n} - |E(G)|$$
$$\implies 4|E(G)|^2 - 2n|E(G)| - n^2(n-1) \leq 0.$$

Solving this quadratic yields the following bound.

Theorem 3.20. *For any n,*

$$\mathrm{ex}(n, C_4) \leq \frac{1}{4}\left(\sqrt{4n-3}+1\right)n.$$

This bound turns out to be sharp. Some of the standard examples include "incidence graphs" of points and lines of finite projective planes.[3] Let's consider the following somewhat simpler example, borrowed from these notes [100].

Example 3.21. Given a prime p, let G_p be the graph with vertex set $\mathbb{Z}_p \times \mathbb{Z}_p$ where (x, y) is adjacent to (x', y') if $x + x' = yy'$ and $(x, y) \neq (x', y')$. For any (x, y) and any choice of x', there is a unique y' for which the equation is satisfied. Thus, each (x, y) has degree either $p - 1$ or p, depending on whether or not the equation is satisfied with $(x', y') = (x, y)$. Therefore, the number of edges is at least

$$\frac{1}{2}p^2(p-1) = \Theta(n^{3/2})$$

where $n = p^2$ is the number of vertices.

Now, let us show that G_p does not have any 4-cycles. Let (x, y) be a vertex and let (x_1, y_1) and (x_2, y_2) be two distinct neighbours of it. Then

$$x + x_1 = yy_1$$

and

$$x + x_2 = yy_2.$$

Subtracting these equations, we get

$$x_1 - x_2 = y(y_1 - y_2).$$

Thus, given the choice of x_1, x_2, y_1 and y_2, provided that $(x_1, y_1) \neq (x_2, y_2)$, the value of y is uniquely determined from this equation. Taking the equation $x + x_1 = yy_1$ then uniquely determines x. Therefore, (x, y) is the unique common neighbour of (x_1, y_1) and (x_2, y_2), and so G_p does not contain a 4-cycle.

3.5 Better Bounds for Even Cycles

In the previous section, we proved an upper bound which holds for any bipartite graph. In this section, we consider the special case of even cycles C_{2k} for $k \geq 2$. In this case, since cycles are rather sparse (and, therefore, are easier to find in a graph), we expect the upper bounds to be much lower than for complete bipartite graphs. The following result was originally proved by Bondy and Simonovits [35].

[3]Finite projective places are discussed in Math 322 at UVic [192, Chapter 4].

Figure 3.11: “Breadth-first search.”

Theorem 3.22 (Bondy and Simonovits [35])**.** *For every $k \geq 2$,*

$$\mathrm{ex}(n, C_{2k}) = O\left(n^{1+\frac{1}{k}}\right)$$

where the constant factor depends on k.

The proof of this theorem is completely elementary, but is perhaps a bit long and involved. Therefore, for the purposes of these notes, we will prove a weaker variant of it. Given an integer n and set $\mathcal{H}$ of graphs, let $\mathrm{ex}(n, \mathcal{H})$ be the maximum number of edges in a graph on n vertices which does not contain a copy of any graph $H \in \mathcal{H}$. We prove the following.

Theorem 3.23. *For any $k \geq 2$ and $n \geq 1$,*

$$\mathrm{ex}(n, \{C_4, C_6, \ldots, C_{2k}\}) \leq 2n\left(n^{1/k} + 1\right).$$

As a first step in the proof of Theorem 3.23, we will “throw away” a bunch of the graph G to obtain a bipartite graph with large minimum degree; the utility of this will be seen later on. The next two lemmas are useful for doing this.

Lemma 3.24. *Every graph G has a bipartite subgraph with at least $|E(G)|/2$ edges.*

Proof. Let A be a subset of $V(G)$ obtained by including each vertex of G in A with probability $1/2$, independently of one another, and then delete all edges with both endpoints in A or both endpoints in $V(G) \setminus A$. This procedure always produces a bipartite subgraph of G. For each edge uv of G, the probability that uv survives this procedure is

$$\mathbb{P}(u \in A \text{ and } v \in V(G) \setminus A) + \mathbb{P}(u \in V(G) \setminus A \text{ and } v \in A) = \frac{1}{2}.$$

Therefore, the expected number of edges in the random bipartite subgraph produced is $|E(G)|/2$, and so there must be a bipartite subgraph of G with at least this many edges. □

Lemma 3.25. *Every graph G has a subgraph of minimum degree at least $|E(G)|/|V(G)|$.*

Proof. The proof is similar to that of Lemma 3.16, but the analysis is simpler. We proceed by induction on $|V(G)|$, where the case $|V(G)| = 1$ is trivial. Now, for $|V(G)| \geq 2$, let x be a vertex of minimum degree in G. If its degree is at least $|E(G)|/|V(G)|$, then we are done. So, suppose that it is less than this and let $G' := G - x$.

By induction, there is a subgraph G'' of G' with minimum degree at least $|E(G')|/|V(G')|$. To complete the proof, all we need to do is to show that $|E(G')|/|V(G')| \geq |E(G)|/|V(G)|$. Indeed, we have

$$\frac{|E(G')|}{|V(G')|} = \frac{|E(G)| - d(x)}{|V(G)| - 1} > \frac{|E(G)| - |E(G)|/|V(G)|}{|V(G)| - 1} = \frac{|E(G)|}{|V(G)|}$$

and so we are done. □

We present a proof of Theorem 3.23.

Proof of Theorem 3.23. Let G be a graph with n vertices and at least $2n(n^{1/k} + 1)$ edges and suppose, to the contrary, that G has no cycle of length $\{4, 6, \ldots, 2k\}$. Using Lemmas 3.24 and 3.25 (in that order), we get a bipartite subgraph G' of G such that

$$\delta(G') \geq \frac{|E(G)|}{2|V(G)|} \geq n^{1/k} + 1.$$

Now, we do a simple "breadth-first search" argument. Let v_0 be any vertex of G' and, for $i \geq 0$, let L_i be the set of vertices at distance exactly i from v_0 in G'. That is,

$$L_0 = \{v_0\},$$

$$L_1 = N_{G'}(v_0),$$

$$L_2 = \bigcup_{u \in L_1} N_{G'}(u) \setminus L_0,$$

and so on. Since G' is bipartite, there are no edges within any of the sets L_i. Also, for $i \leq k-1$, if there are distinct vertices u and v in L_i which have a common neighbour w in L_{i+1}, then, by following the paths form u and v back to v_0 which travel through $L_{i-1}, L_{i-2}, \ldots, L_0$, we eventually reach a common "ancestor" of u and v. However, this gives us an even cycle of length at most $2k$ in G', which is a contradiction. Thus, for $i \leq k-1$, no two vertices of L_i have a common neighbour in L_{i+1}. Thus, each vertex in L_{i+1} has a unique neighbour in L_i. From this, we get that

$$|L_0| = 1,$$

$$|L_1| \geq n^{1/k} + 1$$

and, for $2 \leq i \leq k$,

$$|L_i| \geq |L_{i-1}| n^{1/k}.$$

However, this tells us that $|L_k| \geq \left(n^{1/k}\right)^{k-1} \left(n^{1/k} + 1\right) > n$ which is a contradiction. □

3.6 Exercises

3.1 Show that, if G is a triangle-free graph, then $\alpha(G) \geq \Delta(G)$, where

$$\Delta(G) := \max\{d(v) : v \in V(G)\}$$

and

$$\alpha(G) := \max\{|S| : S \text{ is an independent set in } G\}.$$

3.2 (a) Prove that, for every graph G,

$$\alpha(G) \geq \sum_{v \in V(G)} \frac{1}{d(v)+1}.$$

(b) Use part (a) to prove Turán's Theorem.

3.3 For $r \geq 2$, suppose that G is a graph which does not contain K_r. Prove that there exists a graph G' with $V(G') = V(G)$ such that $\chi(G') \leq r-1$ and

$$d_{G'}(v) \geq d_G(v)$$

for all $v \in V(G)$. Deduce Turán's Theorem.

Hint: Consider a vertex of maximum degree. Induct on r.

3.4 A *tournament* is a complete graph in which every edge is oriented toward one of its endpoints. A k-cycle in a tournament is a sequence of k vertices $(v_0, v_1, \ldots, v_{k-1})$ such that the edge $v_i v_{i+1}$ is oriented from v_i to v_{i+1} for all $0 \leq i \leq k-1$ (viewing indices modulo k). Prove that every tournament on n vertices contains at most $\frac{1}{4}\binom{n+1}{3}$ 3-cycles.

3.5 A tournament is said to be *transitive* if the vertices can be ordered $v_1, v_2, \ldots, v_n$ such that, if $i < j$, then the edge $v_i v_j$ is oriented from v_i to v_j. Prove that every tournament on n vertices contains a transitive tournament with at least $\lfloor \log_2(n) \rfloor + 1$ vertices.

3.6 Given n irrational numbers $x_1, \ldots, x_n$, let $f(x_1, \ldots, x_n)$ be the number of unordered pairs $\{x_i, x_j\}$ such that $x_i + x_j$ is rational. Determine the maximum value of $f(x_1, \ldots, x_n)$ among all choices of $x_1, \ldots, x_n \in \mathbb{R} \setminus \mathbb{Q}$.

3.7 In the proof of Mantel's Theorem, we showed that $\sum_{v \in V(G)} d(v)^2 \leq |E(G)||V(G)|$ for any triangle-free graph G. Prove that, for any graph G (with or without triangles),

$$\sum_{v \in V(G)} d(v)^2 \leq 2|E(G)|(|V(G)|-1).$$

For which graphs G is this inequality tight?

3.8 A *permutation* of $[n]$ is a bijection $\pi : [n] \to [n]$. We often express a permutation π as a sequence $(\pi(1), \ldots, \pi(n))$. For example, $(1,3,2)$ is a permutation of [3]. The *displacement* of a permutation π of $[n]$ is defined to be $\sum_{i=1}^{n} |\pi(i) - i|$. Prove that every permutation π of $[n]$ has displacement at most $n^2/2$.

3.9 Given a permutation σ of $[k]$ and a permutation π of $[n]$ where $k \leq n$, a *copy* of σ in π is a subsequence $(a(1), \ldots, a(k))$ of $(\pi(1), \ldots, \pi(n))$ such that $a(i) < a(j)$ if and only if $\sigma(i) < \sigma(j)$ for $1 \leq i, j \leq k$. Let $\#(\sigma, \pi)$ denote the number of copies of σ in π and define $\#(\sigma, n)$ to be the maximum of $\#(\sigma, \pi)$ over all permutations π of $[n]$. The *packing density* of σ is

$$p(\sigma) := \lim_{n \to \infty} \frac{\#(\sigma, n)}{\binom{n}{k}}.$$

(a) Prove that $p(\sigma)$ is well-defined (i.e. that the limit exists).

(b) By coming up with an explicit construction of permutations $\pi_1, \pi_2, \ldots$ whose lengths tend to infinity, prove a lower bound on $p(1, 3, 2)$. Try to make this lower bound as large as you can.

(c) (Not easy!) Prove that your construction from the previous part is optimal.

3.10 Let M_k be the graph consisting of a matching with k edges. Prove that

$$\mathrm{ex}(n, M_k) = \binom{k-1}{2} + (k-1)(n-k+1)$$

for all $n \geq 3k$.

3.11 Prove Claim 3.6.

3.12 (a) Let $\alpha \geq 0$ and let $x_1, \ldots, x_n$ be non-negative real numbers such that $\sum_{i=1}^{n} x_i = \alpha$. Suppose that $0 < \varepsilon \leq \frac{x_j - x_i}{2}$ for some $1 \leq i, j \leq n$. For $1 \leq s \leq n$, define

$$y_s := \begin{cases} x_s & \text{if } s \notin \{i, j\} \\ x_s + \varepsilon & \text{if } s = i, \\ x_s - \varepsilon & \text{if } s = j. \end{cases}$$

Show that $\sum_{s \neq t} x_s x_t < \sum_{s \neq t} y_s y_t$.

(b) Using part (a), or otherwise, argue that the maximum number of edges in a complete $(r-1)$-partite graph with $n \geq r$ vertices is attained with all parts are of cardinality $\lfloor n/(r-1) \rfloor$ or $\lceil n/(r-1) \rceil$

3.13 Given graphs F and H, let $\mathrm{ex}(F, H)$ be the maximum number of edges in a subgraph of F which does not contain a copy of H. Let Q_d be the *d-dimensional hypercube*; i.e. the graph with vertex set $\{0, 1\}^d$ in which two vertices are adjacent if they differ in a unique coordinate. Prove that

$$\mathrm{ex}(Q_d, C_4) \geq |E(Q_d)|/2.$$

3.14 Suppose that you have a TV remote control and a box of 8 batteries. The remote requires 2 charged batteries in order to function—i.e. if one or more of the batteries is dead, then it will not work. You know that 4 of the batteries in the box are charged and 4 are dead, but you do not know which are which. How many pairs of batteries do you need to test in order to guarantee that one of the tests is successful?

3.15 Compute $\mathrm{ex}(n, C_5)$ for $n \in \{1, 2, 3, 4, 5, 6\}$.

3.16 (a) Show that, for any graph H, the sequence $\mathrm{ex}(n,H)/\binom{n}{2}$ for $n \geq 2$ is non-increasing.

(b) Prove Proposition 3.14.

3.17 For each n, let $g(n)$ be the maximum number of edges in a graph G on n vertices such that the edges of G can be coloured in two colours in such a way that there are no monochromatic triangles.

(a) Prove that $\frac{g(n)}{\binom{n}{2}}$ converges as $n \to \infty$.

(b) Determine $\lim_{n\to\infty} \frac{g(n)}{\binom{n}{2}}$.

3.18 Show that there exists a constant $C > 0$ and a graph H such that $\chi(H) = 3$ and, for every n,

$$\mathrm{ex}(n, H) \geq \frac{n^2}{4} + Cn^{3/2}.$$

Hint 1: Example 3.21 may be helpful.

Hint 2: For every integer $m \geq 1$, there is a prime between m and $2m$ (this is known as Bertrand's Postulate).

3.19 For positive integers n, s, t, let $z(n, s, t)$ be the maximum number of edges in a subgraph of $K_{n,n}$ without a copy of $K_{s,t}$. Prove that $2\,\mathrm{ex}(n, K_{s,t}) \leq z(n, s, t) \leq \mathrm{ex}(2n, K_{s,t})$.

3.20 Let p be a prime and let G_p be a graph with vertex set

$$V(G_p) := \{(x, y, z) \in \mathbb{Z}_p^3 : x \neq y, x \neq z, y \neq z\}$$

where (x_1, y_1, z_1) is adjacent to (x_2, y_2, z_2) if and only if

$$(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 \equiv 1 \bmod p.$$

(a) Determine $|V(G_p)|$ as a function of p for all p.

(b) Prove that every vertex in G_p has degree $p^2 - p$.

(c) Show that G_p does not contain a copy of $K_{3,3}$.

(d) Conclude that $\mathrm{ex}(K_{3,3}, n) = \Theta(n^{5/3})$. (Hint: Use Bertrand's Postulate; see the second hint to Exercise 3.18).

3.21 Determine $\mathrm{ex}(n, C_n)$ for all $n \geq 3$.

3.22 Prove that, for each $k \geq 2$ and $n \geq 1$,

$$\mathrm{ex}(n, \{C_3, C_4, C_5, \ldots, C_{2k}\}) \leq n\left(n^{1/k} + 1\right).$$

3.23 Determine $\mathrm{ex}\left(n, \{C_{2k} : k \geq 2\}\right)$ for every n.

3.24 A *tree* is a connected graph with no cycles. Show that, for every tree T, there exist a constant $c = c(T) \geq 0$ such that $\mathrm{ex}(n, T) \leq cn$ for all n.

3.25 Determine $\mathrm{ex}(n, P_k)$ for every n and k.

3.26 A set $A \subseteq \mathbb{N}$ is called a *Sidon set* if $a, b, c, d \in A$ such that $a + b = c + d$, then $\{a, b\} = \{c, d\}$.

(a) Prove that if A is a Sidon set such that $A \subseteq [n]$, then

$$|A| + \binom{|A|}{2} \leq 2n - 1.$$

Conclude that $|A| < 2\sqrt{n}$.

(b) Prove that, if $A \subseteq [n]$ is a Sidon set and $|A| < n^{1/3}$, then there exists $x \in [n] \setminus A$ such that $A \cup \{x\}$ is a Sidon set.

(c) Let $A \subseteq [n]$ be a Sidon set. Let G be a graph on vertex set $\{u_1, \ldots, u_n\} \cup \{v_1, \ldots, v_{2n}\}$ where $u_i v_j \in E(G)$ if there exists $a \in A$ such that $i + a = j$. Show that G does not contain any 4-cycles.

3.27 Let G be a graph with vertex set $\{v_1, \ldots, v_n\}$ and let $x_1, \ldots, x_n$ be n real variables. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Define

$$f(x) = \sum_{v_i v_j \in E(G)} x_i x_j.$$

Let $S = \{x \in [0, 1]^n : x_1 + \cdots + x_n = 1\}$ and define $\rho := \max_{x \in S} f(x)$.

(a) Prove that the maximum of $f(x)$ for $x \in S$ is attained when the non-zero coordinates of x correspond to the vertices of a clique in G. In other words, prove that there exists a clique $C \subseteq \{v_1, \ldots, v_n\}$ and a vector $x \in S$ such that $x_i = 0$ for all $v_i \notin S$ and $f(x) = \rho$.

(b) Prove that $\rho = \frac{1}{2}\left(1 - \frac{1}{r}\right)$ where r is the number of vertices in the largest clique in G.

(c) Prove that there exists a vector $x \in S$ such that $x_i > 0$ for all i and $f(x) = \rho$ if and only if G is a complete multipartite graph.

(d) Use parts (b) and (c) to prove Turán's Theorem by induction on $n - \lfloor n/r \rfloor$.

3.28 Let S be a set of n points in a circular region with radius 1. Prove that the number of pairs $x, y \in S$ such that the Euclidean distance from x to y is at least $\sqrt{2}$ is at most $n^2/3$.

3.29 Let $f(n)$ be the minimum size of a family $\mathcal{F}$ of subsets of $[n]$ such that, for every set $S \subseteq [n]$ with $|S| \leq 3$, there exists $A, B \in \mathcal{F}$ such that $A \cup B = S$ (note: we allow the case $A = B$). Prove that $f(n) = 1 + n + \binom{n}{2} - \left\lfloor \frac{n^2}{4} \right\rfloor$.

3.30 An *r-uniform hypergraph* is a pair (V, E) such that $E \subseteq \binom{V}{r}$. The elements of V are called *vertices* and elements of E are called *hyperedges*. An r-uniform hypergraph $H = (V, E)$ is *r-partite* if its vertices can be partitioned into r sets $A_1, \ldots, A_r$ such that $|e \cap A_i| = 1$ for every $e \in E$ and $1 \leq i \leq r$. Prove that every r-uniform hypergraph $H = (V, E)$ has an r-partite subhypergraph with at least $\frac{r!}{r^r}|E|$ hyperedges.

3.31 For $r \geq 2$ and $n \geq 1$, the *complete* r-uniform hypergraph on n vertices, denoted $K_n^{(r)}$, is the hypergraph with vertex set $[n]$ and edge set $\binom{[n]}{r}$. Here, we analyze a well-known construction in the area of hypergraph Turán problems. Let n be divisible by 3 and let V_0, V_1, V_2 be disjoint sets of cardinality $n/3$. Define $V := V_0 \cup V_1 \cup V_2$ and let H be a 3-uniform hypergraph with vertex set V such that $e \in E(H)$ if and only if either (i) $|e \cap V_i| = 1$ for all $i \in \{0,1,2\}$ or (ii) there exists $i \in \{0,1,2\}$ such that $|e \cap V_i| = 2$ and $|e \cap V_{i+1}| = 1$, where the indices are viewed modulo 3.

(a) Prove that $|E(H)| \sim \frac{5}{9}\binom{n}{3}$.

(b) Prove that H does not contain a copy of $K_4^{(3)}$.

3.32 Let $K_4^{(3)}$ be the complete 3-uniform hypergraph on 4 vertices, as in Exercise 3.31. Our goal is to show that if H is a 3-uniform hypergraph on n vertices with no copy of $K_4^{(3)}$, then $|E(H)| \leq \frac{2n-3}{9}\binom{n}{2}$.

(a) For distinct $x, y \in V(H)$, let $N(x,y) = \{z \in V(H) : \{x,y,z\} \in E(H)\}$ and $d(x,y) = |N(x,y)|$. Prove that if $\{u,v,w\} \in E(H)$, then $N(u) \cap N(v) \cap N(w) = \emptyset$. Use this to show that, if $\{u,v,w\} \in E(H)$, then $d(u) + d(y) + d(z) \leq 2n - 3$.

(b) By summing over all $e \in E(H)$, prove that

$$\sum_{\{x,y\} \in \binom{V(H)}{2}} d(x,y)^2 \leq (2n-3)|E(H)|.$$

(c) Prove that the left side of the inequality in the previous part of the question is at least $\frac{9|E(H)|^2}{\binom{n}{2}}$. Conclude that $|E(H)| \leq \frac{2n-3}{9}\binom{n}{2}$.

3.33 For $r \geq 2$, $k \geq 1$ and $m \geq 1$, let $K_k^{(r)}(m)$ be an r-uniform hypergraph whose vertices are partitioned into k classes $A_1, A_2, \ldots, A_k$, each of cardinality m, and whose hyperedges are precisely the sets e of cardinality three such that $|e \cap A_i| \leq 1$ for $1 \leq i \leq n$. Given an r-uniform hypergraph H and $n \geq 1$, let $\mathrm{ex}(n, H)$ be the maximum number of hyperedges in an r-uniform hypergraph on n vertices which does not contain a copy of H.

(a) Prove that

$$\mathrm{ex}\left(n, K_r^{(r)}(m)\right) = O\left(n^{r - \frac{1}{m^{r-1}}}\right).$$

(b) Let $K_r(m)$ denote the complete r-partite graph with parts of size m. Using part (a), prove that, for every $\varepsilon > 0$, there exists $\delta > 0$ such that if G is a graph containing at least $\varepsilon\binom{n}{r}$ copies of K_r, then G has at least $\delta\binom{n}{rm}$ copies of $K_r(m)$.

3.34 Define a notion of *Turán density* for k-uniform hypergraphs that is analogous to Definition 3.13. What do Exercises 3.31, 3.32 and 3.33 say about the Turán density of $K_4^{(3)}$ and $K_r^{(r)}(m)$?

3.7 Challenge Problems

3.1* For $n \geq 1$, let X and Y be independent identically distributed random variables in $\mathbb{R}^n$. Prove that, for every $x \geq 0$,

$$\mathbb{P}(|X+Y| \geq x) \geq \frac{1}{2}\mathbb{P}(|X| \geq x)^2.$$

3.2* Given graphs F and H, let $\mathrm{ex}(F, H)$ be the maximum number of edges in a subgraph of F which does not contain a copy of H. Let Q_d be the *d-dimensional hypercube*; i.e. the graph with vertex set $\{0,1\}^d$ in which two vertices are adjacent if they differ in a unique coordinate. Prove that

$$\mathrm{ex}(Q_d, C_6) \geq \frac{1}{3}|E(Q_d)|.$$

3.3* The *packing density* of a permutation σ is defined in Exercise 3.9. Find the packing density of *any* permutation σ of length $k \geq 4$ such that σ is not the increasing permutation $(1, 2, \ldots, k)$ or the decreasing permutation $(k, k-1, \ldots, 1)$. (You get to choose k and σ).

Chapter 4

Beyond or Below the Extremal Number

In this chapter, we focus on more detailed and subtle questions than computing or estimating extremal numbers. In particular, we will consider the following types of questions:

(1) Suppose that G is a graph on n vertices such that $|E(G)| > \mathrm{ex}(n, H)$; i.e. that G is *supersaturated*. How many copies of H must G contain?

(2) Suppose that G is an H-free graph with n vertices such that $|E(G)|$ is "almost as large as" $\mathrm{ex}(n, H)$. Does it follow that the "structure" of G is "close" to that of an extremal example? That is, are the extremal graphs *stable*?

With regards to Question (1), we will begin by proving tight results for some specific graphs H, including K_3 and even cycles. After this, we will prove a "qualitative" result for general graphs which says that if the number of edges in G is larger than the extremal number by εn^2, then, as n grows, the number of copies of H in G grows at a rate of $\Omega\left(n^{|V(H)|}\right)$. We will explore Question (2) only in the case that H is a complete graph.

4.1 Supersaturation for Triangles

Our next aim is to give a very nice argument of Goodman [122] which gives a nearly-optimal supersaturation result in the case that H is a triangle. This can be thought of as a quantitative strengthening of Mantel's Theorem for graphs with more than $n^2/4$ edges.

Theorem 4.1 (The Goodman Bound [122])**.** *Let G be a graph and let $p = 2|E(G)|/|V(G)|^2$. The number of triangles in G is at least*

$$p(2p-1)|V(G)|^3/6.$$

Proof. We follow an argument similar to the proof of Mantel's Theorem. Let t be the number of triangles in G and, for each edge $uv \in E(G)$, let t_{uv} be the number of triangles which contain uv. Of course, t_{uv} is precisely $|N(u) \cap N(v)|$. So, for any $uv \in E(G)$,

$$d(u) + d(v) = |N(u) \setminus N(v)| + |N(v) \setminus N(u)| + 2|N(u) \cap N(v)|$$

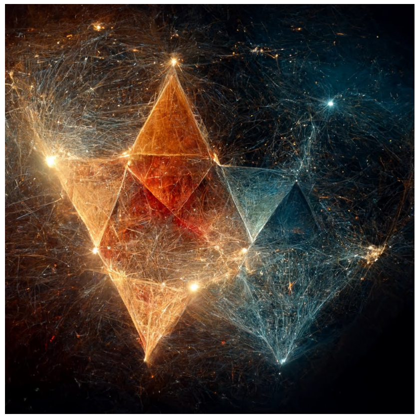


Figure 4.1: “Triangles in a graph.”

$$\leq |V(G)| + t_{uv}.$$

By Lemma 3.3,

$$\sum_{v \in V(G)} d(v)^2 = \sum_{uv \in E(G)} (d(u) + d(v)) \leq |E(G)||V(G)| + \sum_{uv \in E(G)} t_{uv}$$
$$= |E(G)||V(G)| + 3t$$

(since the sum $\sum_{uv \in E(G)} t_{uv}$ counts every triangle thrice). Combining this with the handshaking lemma and Corollary C.7,

$$4|E(G)|^2 = \left(\sum_{v \in V(G)} d(v) \right)^2 \leq |V(G)| \sum_{v \in V(G)} d(v)^2$$
$$\leq |V(G)|^2 |E(G)| + 3|V(G)|t.$$

Thus, by definition of p, we have

$$t \geq \frac{4|E(G)|^2}{3|V(G)|} - \frac{|V(G)||E(G)|}{3} = \frac{p^2|V(G)|^3}{3} - \frac{p|V(G)|^3}{6}.$$

This completes the proof. □

Note that the quadratic $p(2p-1)$ is non-negative whenever $p \geq 1/2$. Therefore, the Goodman Bound implies Mantel's Theorem.

4.2 Supersaturation for Quadrangles

Theorem 3.15 (or Theorem 3.18) implies that the Turán density of any bipartite graph is zero, and so if G is a graph with $\Omega(n^2)$ edges and sufficiently many vertices, then it must contain every small bipartite graph as a subgraph. Our focus in this section is on obtaining a sharp bounds on the

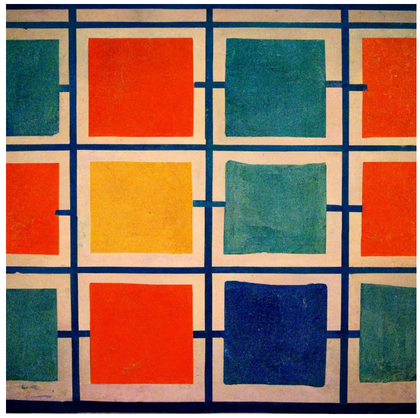


Figure 4.2: “Supersaturated quadrangles.”

number of copies of C_4 in a graph with $\Omega(n^2)$ edges, akin to the Goodman Bound proved in the previous section. Rather than counting “copies” of C_4 in the way that we have been so far, it will be convenient to instead count “homomorphisms” of C_4. Consider the following definition.

Definition 4.2. Given graphs H and G, a *homomorphism* from H to G is a function $f : V(H) \to V(G)$ such that $f(u)f(v) \in E(G)$ whenever $uv \in E(H)$.

The theory of graph homomorphisms has a long and fascinating history [133]; some of the key contributors to this theory include UVic’s very own Gary MacGillivray, Jing Huang and Rick Brewster. The main focus of this area has been on the case where the domain graph H is “large” and one is trying to map it into a small codomain graph H. We will be taking the opposite perspective; that is, we think of H as being a small fixed graph, such as a 4-cycle, and G as being a graph on n vertices for some large n.

Definition 4.3. Let $\mathrm{Hom}(H, G)$ be the set of all homomorphisms from H to G and define $\hom(H, G) = |\mathrm{Hom}(H, G)|$.

Definition 4.4. Given graphs H and G, the *homomorphism density* of G in H is the quantity $t(H, G)$ defined by

$$t(H, G) := \frac{\hom(H, G)}{|V(G)|^{|V(H)|}}.$$

Note that $|V(G)|^{|V(H)|}$ is nothing more than the number of functions from $V(H)$ to $V(H)$; thus, $t(H, G)$ can be equivalently interpreted as the probability that a random function $f : V(H) \to V(G)$ is a homomorphism. It is time for a few examples.

Example 4.5. For any graph G, a homomorphism from K_2 to G is nothing more than a mapping $f : V(K_2) \to V(G)$ such that the two vertices in the image of f are adjacent. Given any edge uv of G, there are precisely two homomorphisms from K_2 to G whose image is $\{u, v\}$. Therefore,

$$t(K_2, G) = \frac{2|E(G)|}{|V(G)|^2}.$$

Looking back to the statement of Theorem 4.1, we see that $t(K_2, G)$ is precisely the value of p in the statement.

Example 4.6. Similarly to Example 4.5, for any graph G, the number of homomorphisms from K_3 to G is 6 times the number of triangles in G. Thus,

$$t(K_3, G) = \frac{6(\text{\# of triangles in } G)}{|V(G)|^3}.$$

Observation 4.7. *Combining Examples 4.5 and 4.6, we see that Theorem 4.1 can be restated rather elegantly as*

$$t(K_3, G) \geq t(K_2, G) \cdot (2t(K_2, G) - 1)$$

for every graph G.

Let us briefly discuss how the homomorphism density $t(H, G)$ relates to counting "copies" of H in G. On one hand, every copy of H in G gives us at least one homomorphism from H to G by simply mapping H to that copy. The other direction is not quite so immediate. Of course, an injective homomorphism from H to G gives us a copy of H in G, but not all homomorphisms are injective. However, the number of non-injective homomorphisms from H to G is always $O\left(|V(G)|^{|V(H)|-1}\right)$. So, if the number of vertices in G is large and $\hom(H, G) \gg |V(G)|^{|V(H)|-1}$, then the contribution of non-injective homomorphisms to $t(H, G)$ is negligible.

Recall from Theorem 3.20 and Example 3.21 that $ex(n, C_4) = \Theta(n^{3/2})$. The corresponding supersaturation question asks: how many 4-cycles must appear in a graph in which the number of edges is significantly greater than $n^{3/2}$? For example, what about a graph with $\Theta(n^2)$ edges? The following result provides a supersaturation bound of this type.

Theorem 4.8 (Sidorenko [239])**.** *For any graph G,*

$$t(C_4, G) \geq t(K_2, G)^4.$$

Proof. We start by proving an analogous bound for the 3-vertex path P_3. The number of homomorphisms from P_3 to G is precisely

$$\sum_{v \in V(G)} d(v)^2$$

which, by Corollary C.7, is at least

$$\frac{1}{|V(G)|} \left(\sum_{v \in V(G)} d(v) \right)^2 = \frac{4|E(G)|^2}{|V(G)|}.$$

Thus, dividing both sides by $|V(G)|^3$ and recalling Example 4.5 yields

$$t(P_3, G) \geq t(K_2, G)^2. \tag{4.9}$$

We now apply a similar argument to turn this into a lower bound on $t(C_4, G)$. The number of homomorphisms from C_4 to G is precisely

$$\sum_{(u,v) \in V(G)^2} |N(u) \cap N(v)|^2$$

which, by Corollary C.7, is at least

$$\frac{1}{|V(G)|^2}\left(\sum_{(u,v)\in V(G)^2}|N(u)\cap N(v)|\right)^2.$$

Now, the key observation here is that $\sum_{(u,v)\in V(G)^2}|N(u)\cap N(v)|$ is, in fact, also equal to the number of homomorphisms from P_3 to G. Therefore, dividing both sides by $|V(G)|^4$ and applying (4.9) gives us

$$t(C_4,G)\geq \frac{1}{|V(G)|^6}\left(t(P_3,G)|V(G)|^3\right)^2 = t(P_3,G)^2 \geq t(K_2,G)^4.$$

This completes the proof. □

Let us conclude this section by discussing the tightness of the bound in Theorem 4.8. For $p\in(0,1)$ and large n, suppose that we let G be a graph on n vertices in which every edge of K_n is included in G with probability p, independent of all other edges. Then the expected number of edges in G is

$$p\binom{n}{2}$$

and so, for n very large, we expect $t(K_2,G)\approx p$. Now, as we discussed before, if n is large, then the number of non-injective homomorphisms from C_4 to G is $O(n^3)$. The expected number of injective homomorphisms is precisely

$$p^4 n(n-1)(n-2)(n-3)$$

and so we expect $t(C_4,G)\approx p^4$ for large n. Thus, in expectation, the bound in Theorem 4.8 is approximately an equality for this graph. Thus, one way to interpret Theorem 4.8 is that the homomorphism density of C_4 in a graph with a given K_2 density is asymptotically minimized by a random graph of that density.

4.3 Supersaturation for Even Cycles

Our next goal is to prove a generalization of Theorem 4.8 to all even cycles. As it turns out, the proof that we gave for 4-cycles does not easily generalize to all even cycles (c.f. Challenge Problem 4.1*). Instead, we will establish a link between the number of cycles of even length in a graph G and the eigenvalues of its adjacency matrix and apply some basic linear algebra.

Definition 4.10. Let G be a graph with vertices $v_1,\dots,v_n$. The *adjacency matrix* of G, denoted A_G, is an $n\times n$ matrix in which the entry in the ith row and jth column is 1 if $v_iv_j\in E(G)$ and 0 otherwise.

Note that the adjacency matrix of G depends on the labelling of the vertices. So, it is not unique; i.e. a given graph G may have many different adjacency matrices.

The adjacency relation is symmetric; i.e., if u is adjacent to v, then v is adjacent to u. This implies that adjacency matrices are symmetric, in the sense that $A_G^T = A_G$ where A^T is the transpose of a matrix A. All of the statements in the following lemma are usually proven in a standard course on linear algebra. Let $\mathrm{tr}(A)$ denote the *trace* of a square matrix A; i.e. the sum of its diagonal entries. Given a square matrix A, let $A^1=A$ and $A_k = AA^{k-1}$ for $k\geq 2$. Let $\langle\cdot,\cdot\rangle$ be the standard inner product on $\mathbb{R}^n$.

Figure 4.3: "Spectrum."

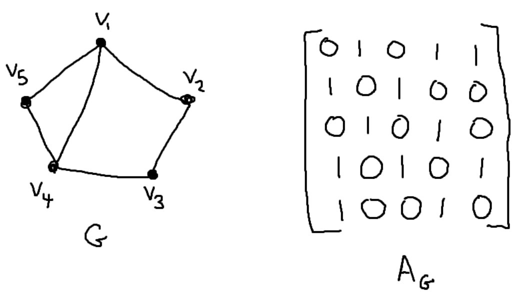


Figure 4.4: A graph G with vertices labelled $v_1, \ldots, v_5$ and the adjacency matrix of G with respect to this labelling.

Lemma 4.11. *Let A be an $n \times n$ matrix over $\mathbb{R}$. Let $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ be the eigenvalues of A.*

(a) If A is symmetric, then $\lambda_1, \ldots, \lambda_n$ are real.

(b) (Min-Max Theorem): If A is symmetric, then $\lambda_1 = \frac{\max_{\vec{v} \neq \vec{0}} \langle \vec{v}, A\vec{v} \rangle}{\langle \vec{v}, \vec{v} \rangle}$ where λ_1 is the largest eigenvalue of A.

(c) For $k \geq 1$, $\lambda_1^k, \ldots, \lambda_n^k$ are the eigenvalues of A^k.

(d) $\mathrm{tr}(A) = \sum_{i=1}^n \lambda_i$.

The key connection between the eigenvalues of A_G and cycles in G is given by the following lemma and the corollary which follows it. For $k \geq 1$, *walk of length k* in a graph is a sequence $w_1, w_2, \ldots, w_{k+1}$ of vertices such that any two consecutive vertices are adjacent.

Lemma 4.12. *For all $k \geq 1$, and for all $1 \leq i, j \leq n$, the (i, j) entry of A_G^k is equal to the number of walks of length k in G which start at v_i and end at v_j.*

Proof. We proceed by induction on k. For $k = 1$, the result is trivial as there is a walk of length 1 from v_i to v_j if and only if they are adjacent.

Now, for $k \geq 2$, the (i,j) entry of A_G^k is the sum over all $1 \leq \ell \leq n$ of the (i,ℓ) entry of A_G^{k-1} multiplied by the (ℓ,j) entry of A_G. Thus, by induction, the ℓth term counts the number of walks of length $k-1$ from v_i to v_ℓ times the number of walks of length 1 from v_ℓ to v_j. That is, it counts walks of length k from v_i to v_j in which the penultimate vertex is v_ℓ. Therefore, the sum of this quantity over all ℓ counts all walks of length k from v_i to v_j and we are done. □

Denote the vertex set of the cycle C_k by $\{u_1,\dots,u_k\}$ where the vertices are listed in the order that they come on the cycle.

Corollary 4.13. *For $k \geq 3$, the ith diagonal entry of A_G^k is equal to the number of homomorphisms f from C_k to G such that $f(u_1) = v_i$.*

Proof. For $k \geq 3$, the number of walks of length k that start and end at v_i is precisely the number of homomorphisms f from C_k to G such that $f(u_1) = v_i$. Therefore, the result follows from Lemma 4.12. □

Corollary 4.14. *If G is a graph on n vertices and $\lambda_1,\dots,\lambda_n$ are the eigenvalues of its adjacency matrix, then, for every $k \geq 3$,*

$$t(C_k,G) = \sum_{i=1}^{n} \left(\frac{\lambda_i}{n}\right)^k.$$

Proof. By Corollary 4.13, we have that

$$\operatorname{tr}(A_G^k) = \hom(C_k,G).$$

The result now follows by parts (c) and (d) of Lemma 4.11 and the fact that $t(C_k,G) = \frac{\hom(C_k,G)}{n^k}$. □

The key thing about even cycles is that any real number raised to an even exponent is non-negative. Therefore, by Lemma 4.11 (a), to get a lower bound on $t(C_k,G)$, it suffices to lower bound the largest eigenvalue. Let's put this observation into practice.

Theorem 4.15 (Sidorenko [239])**.** *Let $k \geq 2$. For every graph G,*

$$t(C_{2k},G) \geq t(K_2,G)^{2k}.$$

Proof. By Corollary 4.14,

$$t(C_{2k},G) = \sum_{i=1}^{n} \left(\frac{\lambda_i}{n}\right)^{2k}.$$

By Lemma 4.11 (a), the eigenvalues of A_G are real. Therefore, since $2k$ is even, we have

$$t(C_{2k},G) \geq (\lambda_1/n)^{2k}.$$

Let $\vec{1} = (1,\dots,1)$ be the all-ones vector in $\mathbb{R}^n$. By Lemma 4.11 (b),

$$\lambda_1 \geq \frac{\langle \vec{1}, A_G\vec{1}\rangle}{\langle \vec{1},\vec{1}\rangle} = \frac{2|E(G)|}{n} = t(K_2,G)n.$$

This completes the proof. □

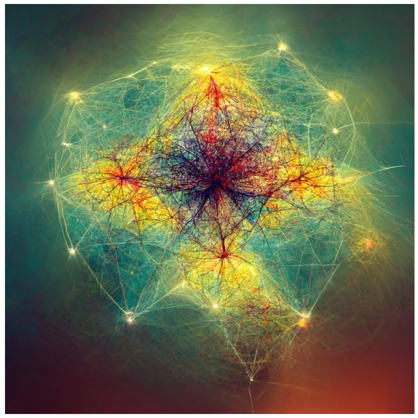


Figure 4.5: "Graph supersaturation."

4.4 Approximate Supersaturation for General Graphs

The following result of Erdős and Simonovits [87] says that, if the number of edges in a graph G on n vertces exceeds the amount given by the extremal density by a quadratic amount, then G contains a positive proportion of the possible copies of H.

Theorem 4.16 (Erdős and Simonovits [87])**.** *For every graph H and $\varepsilon > 0$, there exists $c = c(H, \varepsilon)$ and $n_0 = n_0(H, \varepsilon)$ such that if G is a graph $n \geq n_0$ vertices and at least $(\pi(H) + \varepsilon)\binom{n}{2}$ edges, then the number of copies of H in G is at least $cn^{|V(H)|}$.*

Proof. We assume that $\varepsilon \leq 1-\pi(H)$; otherwise, the theorem holds vacuously, as no such graph can exist. By definition of $\pi(H)$, we may choose n_0 large enough so that $\mathrm{ex}(n, H) \leq \left(\pi(H) + \frac{\varepsilon}{2}\right)\binom{n}{2}$ for all $n \geq n_0$. Now, for $n \geq n_0$, let G be a graph with n vertices which satisfies the hypotheses of the theorem.

By "double counting," we have that

$$\binom{n-2}{n_0-2}|E(G)| = \sum_{\substack{S \subseteq V(G) \\ |S| = n_0}} e(S).$$

This is because every edge of G is contained in precisely $\binom{n-2}{n_0-2}$ sets $S \subseteq V(G)$ of cardinality n_0. Now, say that a subset $S \subseteq V(G)$ of cardinality n_0 is *dense* if $G[S]$ contains more than $\left(\pi(H) + \frac{\varepsilon}{2}\right)\binom{n_0}{2}$ edges and *sparse* otherwise. Let $\mathcal{D}$ and $\mathcal{S}$ be the collection of all dense or sparse sets, respectively. We have

$$\binom{n-2}{n_0-2}|E(G)| = \sum_{S \in \mathcal{D}} e(S) + \sum_{S \in \mathcal{S}} e(S)$$

$$\leq |\mathcal{D}|\binom{n_0}{2} + |\mathcal{S}|\left(\pi(H) + \frac{\varepsilon}{2}\right)\binom{n_0}{2}.$$

If we write $\alpha := |\mathcal{D}|/\binom{n}{n_0}$ and apply the hypothesis on G, then

$$(\pi(H)+\varepsilon)\binom{n-2}{n_0-2}\binom{n}{2} \le \alpha\binom{n}{n_0}\binom{n_0}{2} + (1-\alpha)\left(\pi(H)+\frac{\varepsilon}{2}\right)\binom{n}{n_0}\binom{n_0}{2}.$$

Now, $\binom{n-2}{n_0-2}\binom{n}{2} = \binom{n}{n_0}\binom{n_0}{2}$ (this can be easily shown algebraically, but there is also a simple "combinatorial proof"). Thus, solving for α gives

$$\alpha \ge \frac{\varepsilon}{2\left(1-\pi(H)-\frac{\varepsilon}{2}\right)}.$$

So,

$$|\mathcal{D}| \ge \frac{\varepsilon}{2\left(1-\pi(H)-\frac{\varepsilon}{2}\right)}\binom{n}{n_0} \ge c' n^{n_0}$$

where c' is a constant depending on ε and H only (note that, in some sense, c' also depends on n_0, but that n_0 itself is a function of ε and H; thus, c' can be thought of as depending on ε and H only).

Now, by definition of n_0 and $\mathcal{D}$, for every $S \in \mathcal{D}$, the set $G[S]$ must contain a copy of H. Also, any copy of H is contained in at most $\binom{n-|V(H)|}{n_0-|V(H)|}$ sets of cardinality n_0. Therefore, the number of copies of H in G is at least

$$|\mathcal{D}|/\binom{n-|V(H)|}{n_0-|V(H)|} \ge cn^{|V(H)|}$$

for some constant c which depends on ε and H only. □

Using standard asymptotic notation (see Appendix A), an equivalent restatement of Theorem 4.16 is as follows: *Let H be a graph. If G is a graph with n vertices and $o\left(n^{|V(H)|}\right)$ copies of H, then*

$$|E(G)| \le \pi(H)\binom{n}{2} + o(n^2).$$

4.5 Stability for Turán's Theorem

We close this section by showing that the extremal constructions for Turán's Theorem are "stable;" that is, if G is a K_r-free graph on n vertices in which the number of edges is almost as large as $\mathrm{ex}(n, K_r)$, then we can transform G into a complete $(r-1)$-partite graph by removing and adding a small number of edges.

Theorem 4.17 (Füredi [111])**.** *Let $n \ge r \ge 3$ and $t \ge 1$ and let G be a K_r-free graph on n vertices such that*

$$|E(G)| = \mathrm{ex}(n, K_r) - t.$$

Then there exists $E_1 \subseteq E(G)$ and $E_2 \subseteq E(\overline{G})$ such that $(G - E_1) \cup E_2$ is a complete $(r-1)$-partite graph and

$$|E_1 \cup E_2| \le 3t.$$

Figure 4.6: "Stability."

Proof. What we will prove is that there is a set $E_1 \subseteq E(G)$ of cardinality at most t such that $G - E_1$ has chromatic number at most $r-1$. We claim that this is sufficient to prove the theorem. Indeed, if such a set E_1 exists, then $G - E_1$ is a graph of chromatic number at most $(r-1)$ with at least $\mathrm{ex}(n, K_r) - 2t$ edges. If we let E_2 be the set of all missing edges between vertices in different colour classes, we get that $(G - E_1) \cup E_2$ is a complete $(r-1)$-partite graph. This graph is K_r-free, and so it must have at most $\mathrm{ex}(n, K_r)$ edges; thus, the number of edges in E_2 is at most $2t$.

So, we focus on proving that a set E_1 of the type described above exists. We analyse the so-called Erdős degree majorization algorithm [78]. Initialize

$$V_0^+ := V(G).$$

Now, for $i \geq 1$, if V_{i-1}^+ is non-empty, then let x_i be a vertex of maximum degree in $G[V_{i-1}^+]$. Define

$$V_i^+ := V_{i-1}^+ \cap N(x_i)$$

and

$$V_i := V_{i-1}^+ \setminus N(x_i).$$

By a simple handshaking argument, we have

$$\sum_{v \in V_i} |N(v) \cap V_{i-1}^+| = 2e(V_i) + e(V_i, V_i^+).$$

By our choice of x_i, we have that $|N(v) \cap V_{i-1}^+| \leq |N(x_i) \cap V_{i-1}^+| = |V_i^+|$ for all $v \in V_i$. Combining this with the above equality gives us

$$2e(V_i) + e(V_i, V_i^+) \leq |V_i^+||V_i|. \tag{4.18}$$

Let s be the minimum integer so that $V_s^+ = \emptyset$; i.e. this is when the procedure terminates. Note that we must have $s \leq r-1$ since the vertices $x_1, \ldots, x_s$ form a complete graph. By construction, the sets $V_1, \ldots, V_s$ partition $V(G)$.

Now, to finish the proof, we sum up the inequality (4.18) over $1 \leq i \leq s$. The left side of the resulting inequality counts each edge within any of the sets V_i exactly twice and any edge from V_i

to V_j for $i < j$ exactly once; to see why the latter is true, observe that $V_i^+ = \bigcup_{j=i+1}^s V_j$ for all i. Therefore, in total, the left side is equal to

$$|E(G)| + \sum_{i=1}^{s} e(V_i) = \mathrm{ex}(n, K_r) - t + \sum_{i=1}^{s} e(V_i).$$

The right side precisely counts the number of edges in a complete s-partite graphs whose parts are $V_1, \dots, V_s$, which is at most $\mathrm{ex}(n, K_r)$ (here, we use that $s \leq r - 1$). Therefore,

$$\mathrm{ex}(n, K_r) - t + \sum_{i=1}^{s} e(V_i) \leq \mathrm{ex}(n, K_r)$$

which implies that $\sum_{i=1}^{s} e(V_i) \leq t$. Define E_1 to be the set of all edges which lie within one of the sets V_i. □

Technically, the above result tells us that, if the number of edges in G is close to the number in a Turán graph, then G must be close to "some" complete multipartite graph, but it does not necessarily say that it is close to a Turán graph. However, if we analyse the final step of the proof a little more closely, we see that it actually tells us that the number of edges in E_1 is at most t plus the number of edges in a complete multipartite graph with parts $V_1, \dots, V_s$, minus $\mathrm{ex}(n, K_r)$. Therefore, the number of edges in a complete multipartite graph with parts $V_1, \dots, V_s$ must be at least $\mathrm{ex}(n, K_r)$ minus t. It is now just a tedious calculation to prove that a complete multipartite graph with parts $V_1, \dots, V_s$ can be transformed into a Turán graph by adding or removing a relatively small number of edges; see Exercise 4.6. From this, we get the following:

Theorem 4.19 (Stability for Turán's Theorem)**.** *For every $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that if G is a K_r-free graph on n vertices with*

$$|E(G)| \geq \mathrm{ex}(n, K_r) - \delta n^2,$$

then G can be made into a Turán graph by adding and deleting at most εn^2 edges.

4.6 Exercises

4.1 Prove that, for $k \geq 3$, if G is a graph with n vertices such that every vertex has degree at least $\left(\frac{k-2}{k-1}\right) n + 1$, then every edge of G is contained in a copy of K_k.

(Hint 1: It might help to start with the cases $k = 3$ and $k = 4$ and try to generalize.)

(Hint 2: Turán's Theorem implies that if G has n vertices and more than $\left(\frac{k-2}{k-1}\right) \frac{n^2}{2}$ edges, then it contains a copy of K_k. You may want to use this.)

4.2 (a) Let $k, m \geq 2$ be integers. The *k-colour Ramsey number* of K_m, denoted by $R_k(m)$, is the minimum N such that, in every colouring of the edges of K_n with k colours, there is a monochromatic copy of K_m. Let $r_k(m; n)$ denote the minimum number of monochromatic copies of K_m in a colouring of the edges of K_n with k colours. Prove that

$$r_k(m; n) \geq \binom{n}{R_k(m)} \cdot \binom{n-m}{R_k(m) - m}^{-1}$$

for all $n \geq 1$.

(b) Prove that

$$r_k(m;n) \leq \left(\frac{1}{k}\right)^{\binom{m}{2}-1} \binom{n}{m}$$

for all $n \geq 1$.

(c) Prove that

$$r_k(m;n) \leq (R_{k-1}(m)-1) \cdot \binom{\lceil n/(R_{k-1}(m)-1)\rceil}{m}$$

for all $n \geq 1$.

(d) For the case $k = m = 3$, which of the upper bounds on $r_3(3;n)$ in the previous two parts of the question is better for large n?

(e) Let $k = 3$. Make a conjecture which naturally generalizes the result of the previous part of the question from the case $m = 3$ to all $m \geq 3$.[1] Show that this conjecture would imply an asymptotic upper bound on $R_2(m)$ for all $m \geq 3$ of the form $R_2(m) \leq c^{(1+o(1))m}$ for some $c > 1$. Make c as small as you can.

(f) Do an internet search to find the best known asymptotic upper bound on $R_2(m)$ (other sources may use different notation, e.g. $R(m,m)$, rather than $R_2(m)$). Explain why the bound on $R_2(m)$ that follows from the conjecture in the previous part of this question is better than the best known upper bound when m is very large.

4.3 Let $n \geq 1$ and $1 \leq m \leq \binom{n}{2}$. Describe the graph G which has n vertices and m edges and, subject to this, the maximum number of triangles. Prove it is optimal.

Hint: Think about the collection of triangles as being a subset $\mathcal{T}$ of $\binom{[n]}{3}$. One of the theorems in Chapter 2 may be helpful.

4.4 (a) Show that, if n is even, the graph $K_{n/2,n/2}$ is a tight example for the Goodman Bound in the case that $p = 1/2$. That is, it has the right number of edges and that the number of triangles exactly matches the lower bound in the theorem.

(b) Show that, if n is divisible by 3, then there exists a tight example for the Goodman Bound on n vertices with $p = 2/3$.

(c) For each $k \geq 2$, find a tight example for the Goodman Bound for the case $p = 1 - \frac{1}{k}$.

4.5 The Erdős–Szekeres Theorem[2] says that if r and s are integers and $n \geq (r-1)(s-1)+1$, then any sequence $a_1, \ldots, a_n$ of real numbers contains either an increasing sequence of length r or a decreasing sequence of length s. Prove that, for any $r \geq 2$ and $\varepsilon > 0$, there exists $n_0 = n_0(r, \varepsilon)$ such that if $n \geq n_0$, then every sequence of n real numbers contains at least

$$\left(\frac{((r-1)(r-2))!}{((r-1)^2+1)!} - \varepsilon\right) n^r$$

increasing or decreasing subsequences of length r.

[1] Note: It is currently unknown whether this generalization is true or not.

[2] The Erdős–Szekeres Theorem is covered in Math 322 at UVic [192].

4.6 (a) Let $a_1, \dots, a_k$ be non-negative integers and define $n = \sum_{i=1}^k a_i$. Show that the number of edges in a complete k-partite graph with parts of size $a_1, \dots, a_k$ is equal to

$$\frac{1}{2}\left(n^2 - \sum_{i=1}^{k} a_i^2\right).$$

(b) Assuming (for simplicity) that $r-1$ divides n, use part (a) to prove that, if G is a complete $(r-1)$-partite graph with parts of size $a_1, \dots, a_{r-1}$ such that $\sum_{i=1}^{r-1} a_i = n$, then

$$\operatorname{ex}(n, K_r) - |E(G)| = \frac{1}{2}\sum_{i=1}^{r-1}\left(a_i^2 - \left(\frac{n}{r-1}\right)^2\right).$$

Also, prove that this sum can be rewritten as follows:

$$\sum_{i=1}^{r-1}\left(a_i^2 - \left(\frac{n}{r-1}\right)^2\right) = \sum_{i=1}^{r-1}\left(a_i - \frac{n}{r-1}\right)^2.$$

(c) Using part (b) and Corollary C.7, prove that, if G is a complete $(r-1)$-partite graph with n vertices such that $|E(G)| \geq \operatorname{ex}(n, K_r) - t$, then G can be transformed into a Turán graph by adding or removing at most $n\sqrt{2t/(r-1)}$ edges.

(d) Using Theorem 4.17 and (c), prove the following:

For every $r \geq 3$ and $\varepsilon > 0$ there exists $\delta = \delta(r, \varepsilon) > 0$ and $n_0 = n_0(r, \varepsilon)$ such that if G is a K_r-free graph with $n \geq n_0$ vertices and at least

$$\operatorname{ex}(n, K_r) - \delta n^2$$

edges, then G can be transformed into a Turán graph by adding and removing at most εn^2 edges.

4.7 The goal of this exercise is to lead you through a proof of a lower bound on the sum of the number of triangles in G and the number of triangles in $\overline{G}$. This was proved by Goodman [122].

(a) Suppose that the edges of G are red and the edges of $\overline{G}$ are blue. Argue that the sum $\sum_{v \in V(G)} \binom{d_G(v)}{2}$ counts every triangle with three red edges exactly three times and every triangle with two red edges and one blue edge exactly once.

(b) Make a similar observation about the sum $\sum_{v \in V(G)} \binom{d_{\overline{G}}(v)}{2}$. You don't have to justify it.

(c) Using parts (a) and (b), prove that the number of triangles in G plus the number of triangles in $\overline{G}$ is equal to

$$\frac{1}{2}\left(\sum_{v \in V(G)} \binom{d_G(v)}{2} + \sum_{v \in V(G)} \binom{d_{\overline{G}}(v)}{2} - \binom{n}{3}\right).$$

(d) Optimize the equation from (c) over all choices of degrees to show that the number of triangles in G plus the number of triangles in $\overline{G}$ is always at least

$$\frac{1}{4}\left(\frac{n(n-1)(n-5)}{6}\right) = \frac{1}{4}\binom{n}{3} + O(n^2).$$

4.8 (a) Suppose that n is even and let $G = K_{n/2,n/2}$. What is the sum of the number of triangles in G and the number of triangles in $\overline{G}$?

(b) Suppose that G is a graph on n vertices such that each edge of K_n is included in G with probability $1/2$, independently of all of the other edges. What is the expected value of the sum of the number of triangles in G and the number of triangles in $\overline{G}$?

4.9 Given graphs H_1 and H_2, let $H_1 \sqcup H_2$ denote the disjoint union of H_1 and H_2. Prove that, for any graphs H_1, H_2 and G,

$$t(H_1 \sqcup H_2, G) = t(H_1, G)t(H_2, G).$$

4.10 Given graphs G_1 and G_2, let $G_1 \times G_2$ be the graph with vertex set $V(G_1) \times V(G_2)$ and edge set

$$\{(u, v)(w, x) : uw \in E(G_1) \text{ and } vx \in E(G_2)\}.$$

The graph $G_1 \times G_2$ is called the *tensor* or *categorical* product of G_1 and G_2 (among other names). Prove that, for any graphs H, G_1 and G_2,

$$t(H, G_1 \times G_2) = t(H, G_1)t(H, G_2).$$

4.11 By imitating the proof of Theorem 4.8, prove that $t(K_{4,4}, G) \geq t(C_4, G)^4$ for any graph G. Also, for fixed $p \in (0, 1)$, provide an example to show that this is asymptotically tight for graphs G with $t(C_4, G) = p^4$.

4.12 Let n be an even natural number and let G be a subgraph of $K_{n/2,n/2}$ with $|V(G)| = n$. Prove that

$$t(C_4, G) \geq 2 \cdot t(P_3, G)^2.$$

Hint: Follow the proof of Theorem 4.12 after equation (4.13).

4.13 Prove that, for every $s, t \geq 1$,

$$t(K_{s,t}, G) \geq t(K_2, G)^{st}.$$

Hint: Use Corollary C.6 twice.

4.14 Let B be the 5-vertex "bowtie" graph obtained by gluing two triangles together on a vertex. Prove that $t(B, G) \geq t(K_3, G)^2$ for any graph G.

4.15 Let K_4^- be the graph obtained from K_4 by deleting one edge. Prove that $t(K_4^-, G)t(K_2, G) \geq t(K_3, G)^2$ for every graph G.

4.16 Prove that $t(P_5, G) \geq t(K_2, G)^4$ for every graph G. Generalize the proof to yield $t(P_{2^k+1}, G) \geq t(K_2, G)^{2^k}$ for any $k \geq 1$.

4.17 A *k-edge coloured graph* is a tuple $\vec{H} := (H_1, \ldots, H_k)$ where $H_1, \ldots, H_k$ are graphs with the same vertex set; i.e. $V(H_i) = V(H_j)$ for all $1 \leq i, j \leq k$. We let $V(\vec{H})$ denote $V(H_i)$ for any i. You can visualize a k-edge coloured graph as being k graphs $H_1, \ldots, H_k$ drawn on top of each other, where the edges of H_i are in colour i. If $\vec{H}$ and $\vec{G}$ are k-edge coloured graphs, then $f : V(\vec{H}) \to V(\vec{G})$ is a *homomorphism* if, for any $1 \leq i \leq k$ and $uv \in E(H_i)$, it holds that $f(u)f(v) \in E(G_i)$. Define $\hom(\vec{H}, \vec{G})$ and $t(\vec{H}, \vec{G})$ analogously to the case of graphs.

(a) Let $\vec{H} := (H_1, \dots, H_k)$ be a k-edge coloured graph, let H be a graph with vertex set $V(\vec{H})$ and edge set $\bigcup_{i=1}^k E(H_i)$ and let G be any graph. Let $\vec{G}$ be the k-edge coloured graph $(G_1, \dots, G_k)$ where $G_1 = G_2 = \cdots = G_k = G$. Prove that $t(\vec{H}, \vec{G}) = t(H, G)$.

(b) Let $\vec{C_4} = (H_1, H_2, H_3, H_4)$ be the 4-edge coloured graph on vertex set $\{1, 2, 3, 4\}$ where, for $1 \leq i \leq 4$, the graph H_i contains only the edge $i(i+1)$ where we view the vertices modulo 4 (i.e. $5 \equiv 1$). Prove that, for any 4-coloured graph $\vec{G}$,

$$t(\vec{C_4}, \vec{G})^2 \leq t((H_1, H_2, H_3, H_4), (G_1, G_1, G_3, G_3)) t((H_1, H_2, H_3, H_4), (G_2, G_2, G_4, G_4)).$$

(c) Prove that, for any 4-coloured graph $\vec{G}$,

$$t(\vec{C_4}, \vec{G})^4 \leq t(C_4, G_1) t(C_4, G_2) t(C_4, G_3) t(C_4, G_4).$$

(Hint: The previous two parts of the question are useful).

(d) Let H be a graph with k edges and let $\vec{H} = (H_1, \dots, H_k)$ be a k-edge coloured graph with vertex set $V(H)$ where each edge of H belongs to exacty one of the graphs $H_1, \dots, H_k$. Suppose that

$$t(\vec{H}, \vec{G})^k \leq \prod_{i=1}^k t(H, G_i)$$

for any k-edge coloured graph $\vec{G} = (G_1, \dots, G_k)$. Prove that

$$t(H, G) \geq t(K_2, G)^{|E(H)|}$$

for every graph H. (Hint: Let $G_1 = G$ and let $G_2, \dots, G_k$ be appropriately chosen random graphs).

(e) Deduce Theorem 4.8 from the previous parts of this question.

4.18 Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of the adjacency matrix of a graph G on n vertices. Prove that $\sum_{i=1}^n \lambda_i^2 = 2|E(G)|$.

4.19 (a) Prove that every graph G satisfies

$$t(C_{24}, G) \leq t(C_{42}, G)^{2/5} t(C_{12}, G)^{3/5}.$$

Hint: Apply Corollary 4.14 and Lemma C.5.

(b) Describe a sequence $G_1, G_2, \dots$ of graph such that $\lim_{n \to \infty} t(K_2, G_n) = 1/2$ and

$$\lim_{n \to \infty} \left(t(C_{24}, G_n) - t(C_{42}, G_n)^{2/5} t(C_{12}, G_n)^{3/5} \right) = 0.$$

It is enough to describe a correct construction; you do not need to rigorously prove that it is correct.

4.20 Let $\varepsilon_0, \delta_0 > 0$ and suppose that G is a graph such that

$$t(C_4, G) \leq t(K_2, G)^4 + \varepsilon_0^4$$

and
$$t(C_6, G) \le t(K_2, G)^6 + \delta_0^6.$$
Prove that
$$t(C_5, G) \ge t(K_2, G)^5 - \varepsilon_0^2\delta_0^3.$$
Hint: Apply Corollary 4.14 and Lemma C.5.

4.21 Let G be a graph with n vertices and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of its adjacency matrix indexed so that $\lambda_1 \ge \cdots \ge \lambda_n$. Prove that G is bipartite if and only if its spectrum is *symmetric*; i.e., for every $1 \le i \le n$,
$$\lambda_i = -\lambda_{n-i+1}.$$
Conclude that, if G is a bipartite graph on an odd number of vertices, then the adjacency matrix of G is singular.

4.22 Let G be a graph with n vertices and let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of its adjacency matrix indexed so that $\lambda_1 \ge \cdots \ge \lambda_n$. For $S, T \subseteq V(G)$, let
$$E(S, T) = \{uv \in E(G) : u \in S, v \in T\}$$
and
$$E(S) = E(S, S).$$
Given $S \subseteq V(G)$, let $\overline{S} = V(G) \setminus S$.

(a) Prove that, for every set $S \subseteq V(G)$ and $\alpha, \beta > 0$,
$$\lambda_n \le \frac{2\alpha^2|E(S)| + 2\beta^2|E(\overline{S})| - 2\alpha\beta|E(S, \overline{S})|}{\alpha^2|S| + \beta^2|\overline{S}|}.$$
Hint: Choose an appropriate vector $\vec{v}$ and apply Lemma 4.11 (b) to the matrix $-A_G$.

(b) Let $\varepsilon > 0$. Suppose that G is a graph with n vertices and that $S \subseteq V(G)$ such that $|S| = n/2$ and
$$|E(S, \overline{S})| \ge \frac{|E(G)|}{2} + \frac{\varepsilon n^2}{4}.$$
Prove that
$$t(C_4, G) \ge t(K_2, G)^4 + \varepsilon^4.$$

(c) Recall that the characteristic polynomial of the adjacency matrix of G is $p_{A_G}(x) := \prod_{i=1}^n (x - \lambda_i)$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of A_G. Prove that
$$\ln(p_G(x)) = n\ln(x) - \sum_{r=1}^{\infty} \frac{\hom(C_r, G)}{rx^r}.$$
(Hint: Use the Taylor expansion of the natural logarithm).

4.23 Derive the Erdős–Stone Theorem from Exercise 3.33 (b), Turán's Theorem and the Erdős–Simonovits Supersaturation Theorem.

4.7 Challenge Problems

4.1* Prove that $t(P_4, G) \geq t(K_2, G)^3$ for any graph G. Generalize the proof of this and Exercise 4.16 to yield $t(P_k, G) \geq t(K_2, G)^{k-1}$ for any $k \geq 1$.

4.2* Let Q_d be the d-dimensional hypercube. Prove that $t(Q_d, G) \geq t(K_2, G)^{|E(Q_d)|}$ for every graph G. (Hint: Adapting the idea in Exercise 4.17 should work).

Chapter 5

The Regularity Lemma

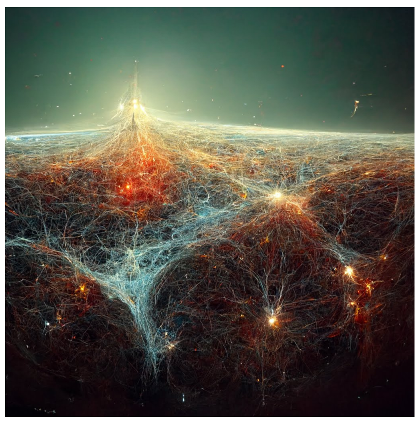


Figure 5.1: “Complex network.”

In the previous two chapters, we have encountered two main types of “extremal” constructions of graphs which minimize or maximize the number of copies of a given graph, under certain constraints:

- Partition the vertex set into a bounded number of “parts.” Join two vertices with an edge if they are in different parts, but not if they are in the same part.

  (Mantel’s Theorem, Turán’s Theorem, the Erdős–Stone Theorem)

- “Throw in” the edges randomly.

  (Theorem 4.8, Exercise 4.8 (b), Exercise 4.11)

The focus of this chapter is on the remarkable fact that the structure of any graph G can be described as a mixture of these two types of constructions, up to small errors; this is known as the Szemerédi Regularity Lemma [253]. Note that, unlike in previous chapters, we are speaking about **any** graph here, with no assumptions on it being an extremal configuration for any particular combinatorial problem. Therefore, the Regularity Lemma is more than just a useful tool for extremal combinatorics; it is a profound statement about the nature of graphs themselves.

5.1 Regular Pairs and Counting

The Regularity Lemma (stated in the next section) roughly says that the vertex set of any graph can be partitioned into a bounded number of parts such that the edges between most pairs of parts are "well behaved." The precise meaning of this is given by Definition 5.2 below.

Definition 5.1. Given a graph G and disjoint non-empty subsets X and Y of $V(G)$, the *density* between X and Y is defined by

$$d(X,Y) := \frac{e(X,Y)}{|X||Y|}.$$

Definition 5.2. Given $\varepsilon > 0$, a graph G and disjoint non-empty subsets X and Y of $V(G)$, say that the pair (X,Y) is *ε-regular* if, for every $A \subseteq X$ and $B \subseteq Y$ such that $|A| \geq \varepsilon|X|$ and $|B| \geq \varepsilon|Y|$,

$$|d(A,B) - d(X,Y)| \leq \varepsilon.$$

In other words, (X,Y) is ε-regular if, for any "large enough" subsets A and B of X and Y, respectively, the density between A and B is the same as the density between X and Y, up to an error of at most ε. Before moving forward, let us make a quick remark and think through a few examples.

Remark 5.3. The condition that A and B are "large enough" is important for making Definition 5.2 meaningful for densities between zero and one. For example, if we allow A and B to be singletons, then $d(A,B)$ is either 0 or 1, depending on whether or not the unique vertex of A is adjacent to the unique vertex of B. Thus, if there was no restriction on $|A|$ and $|B|$, then the only way in which there could exist $0 < \varepsilon < 1/2$ such that (X,Y) is ε-regular is if G contains no edges from X to Y or it contained every edge from X to Y.

Example 5.4. Suppose that X and Y are disjoint sets of vertices and that G contains every edge from X to Y; e.g., if X and Y are two parts of a complete multipartite graph. Then, for every $A \subseteq X$ and $B \subseteq Y$, we have $d(A,B) = 1$. Therefore, the pair (X,Y) is ε-regular for every $\varepsilon > 0$.

Example 5.5. Suppose now that X and Y are disjoint sets of vertices and that every vertex in X has at most d non-neighbours in Y. We claim that, if $\varepsilon > \sqrt{d/|Y|}$, then (X,Y) is ε-regular. For any non-empty sets $A \subseteq X$ and $B \subseteq Y$, we have

$$|A|(|B| - d) \leq e(A,B) \leq |A||B|$$

which implies that

$$(|B| - d)/|B| \leq d(A,B) \leq 1.$$

Thus, if $|B| \geq \varepsilon|Y|$, then

$$|d(A,B) - d(X,Y)| \leq d/|B| \leq d/(\varepsilon|Y|) \leq \varepsilon.$$

Thus, (X,Y) is ε-regular.

Example 5.6. Suppose that X and Y are disjoint sets such that $|X| = |Y|$ and that, for each $x \in X$ and $y \in Y$, the edge xy is included in G with probability $1/2$, independently of all other such edges. For $\varepsilon > 0$, if $|X| = |Y|$ is large with respect to ε, then (X,Y) is ε-regular with high probability.

Here are the details. Given any $A \subseteq X$ and $B \subseteq Y$, by linearity of expectation, the expected number of edges from X to Y is $|A||B|/2$. By the Chernoff Bound (Theorem B.10) and independence, for any $t \geq 0$, the probability that the number of edges deviates from $|A||B|/2$ by at least t is at most $2e^{-2t^2/|A||B|}$. Letting $t = \varepsilon|A||B|/2$, we get that

$$\mathbb{P}(|d(A, B) - 1/2| > \varepsilon/2) \leq 2e^{-\varepsilon^2|A||B|/4}.$$

If each of $|A|$ and $|B|$ is at least $\varepsilon|X| = \varepsilon|Y|$, then this probability is at most $2e^{-\varepsilon^4|X||Y|/4}$. The number of possible choices for (A, B) is at most $2^{|X|}2^{|Y|} = 4^{|X|}$. Thus, the probability that there exists such a pair (A, B) with $|d(A, B) - 1/2| > \varepsilon/2$ goes to zero (very quickly) as $|X|$ goes to infinity. So, with probability close to one, there are no such pairs. Therefore, with probability close to one, (X, Y) is ε-regular.

Example 5.6 is particularly relevant to the way that we will generally "think about" ε-regular pairs. That is, intuitively speaking, if (X, Y) is ε-regular, then we will think of the edges between X and Y as being distributed in a "random-like" way, up to small errors which are bounded in terms of ε.

Now that we are comfortable with the definition of an ε-regular pair, let us turn our attention to their main application; namely, in approximately counting copies of graphs. As a simple special case, let us consider the case of triangles.

Lemma 5.7 (The Counting Lemma For Triangles)**.** *Let G be a graph and let (X_1, X_2, X_3) be a partition of $V(G)$. Let $0 < \varepsilon < 1$ and suppose that, for all $1 \leq i \neq j \leq 3$, the pair (X_i, X_j) is ε-regular and $d(X_i, X_j) > 2\varepsilon$. Then the number of triangles in G with one vertex in each of the sets X_1, X_2 and X_3 is at least*

$$(1 - 2\varepsilon)\left(\prod_{i \neq j}(d(X_i, X_j) - \varepsilon)\right)|X_1||X_2||X_3|.$$

Proof. Given $i \in \{2, 3\}$, say that a vertex $x \in X_1$ is *bad for X_i* if it has fewer than $(d(X_1, X_i) - \varepsilon)|X_i|$ neighbours in X_i. Let $X_{1,i}'$ be the set of all vertices that are bad for X_i. Then, by definition,

$$e(X_{1,i}', X_i) < (d(X_1, X_i) - \varepsilon)\,|X_{1,i}'||X_i|$$

and so $|d(X_{1,i}', X_i) - d(X_1, X_i)| > \varepsilon$. Since (X_1, X_2) is ε-regular, and $|X_i| > \varepsilon|X_i|$, this implies that $|X_{1,i}'| < \varepsilon|X_1|$.

Now, let $Y_1 := X_1 \setminus (X_{1,2}' \cup X_{1,3}')$. As we have shown above, $|Y_1| \geq |X_1| - 2\varepsilon|X_1| = (1 - 2\varepsilon)|X_1$. Note that any $x \in Y_1$ is contained in exactly

$$e(N(x) \cap X_2, N(x) \cap X_3)$$
$$= |N(x) \cap X_2||N(x) \cap X_3| \cdot d(N(x) \cap X_2, N(x) \cap X_3)$$

triangles in G. Since $x \in Y_1$, it has at least $(d(X_1, X_2) - \varepsilon)|X_2| > \varepsilon|X_2|$ neighbours in X_2 and at least $(d(X_1, X_3) - \varepsilon)|X_3| > \varepsilon|X_3|$ in X_3. So, since (X_2, X_3) is ε-regular, the number of triangles containing x is at least

$$(d(X_1, X_2) - \varepsilon)(d(X_1, X_3) - \varepsilon)(d(X_2, X_3) - \varepsilon)|X_2||X_3|.$$

Thus, we are now done by summing the number of triangles containing x over all $x \in Y_1$. $\square$

The main message of Lemma 5.7 is that, if $(X_1, X_2), (X_1, X_3)$ and (X_2, X_3) are all ε-regular, then the number of triangles between these three sets is almost as large as it would be if the edges from X_i to X_j were placed randomly with probability $d(X_i, X_j)$ for all distinct $i, j \in \{1, 2, 3\}$. That is, from the perspective of counting triangles, ε-regular pairs for small ε can be thought of as being "random-like."

Of course, there is nothing particularly special about finding copies of triangles here. A very similar argument works for bounding the number of copies of K_r in a graph whose vertex set is partitioned into r sets $X_1, \ldots, X_r$ such that the pairs (X_i, X_j) with $i \neq j$ are ε-regular. In fact, there is also nothing special about complete graphs. By complementing the edges between pairs (X_i, X_j) and counting appropriately, one can also obtain a lower bound on the number of induced copies of any graph H. Thus, since this holds for all possible subgraphs, by "counting the complement," we can also get upper bounds on the number of induced copies of H. These extensions will be explored in exercises Exercises 5.3 and 5.4.

5.2 The Regularity and Triangle Removal Lemmas

Figure 5.2: "Triangles."

We will now state the Regularity Lemma and highlight its utility with some applications. The proof will be postponed until Section 5.3.

Theorem 5.8 (The Szemerédi Regularity Lemma [253]). *For every $\varepsilon > 0$ and $t \geq 1$ there exists an integer $k_0(\varepsilon, t)$ such that, for every graph G, there exists a partition[1] $X_1, \ldots, X_k$ of $V(G)$ such that*

(a) there are at most εk^2 pairs $(i, j) \in [k]^2$ with $i \neq j$ such that (X_i, X_j) is not ε-regular,

(b) $||X_i| - |V(G)|/k| < 1$ for all $1 \leq i \leq k$ and

(c) $t \leq k \leq k_0(\varepsilon, t)$.

[1]Note that we allow some sets of the partition to be empty (which is necessary in the case $|V(G)| < t$).

Remark 5.9. Without a doubt, the most crucial (and surprising) aspect of the Regularity Lemma is that the upper bound $k_0(\varepsilon, t)$ on the number of sets in the partition depends **only** on the parameters ε and t; in particular, it is independent of the graph G.

As a first application, let us derive the following result, known as the Triangle Removal Lemma. Essentially, this says that a graph with n vertices and $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges. Do not be fooled by its innocent statement; this is an important result with a wide range of consequences.

Lemma 5.10 (The Triangle Removal Lemma [222])**.** *For every $0 < \delta < 1/2$ there exists $\gamma = \gamma(\delta) > 0$ such that if G is a graph with at most $\gamma|V(G)|^3$ triangles, then it can be made triangle-free by removing at most $\delta|V(G)|^2$ edges.*

Proof. We prove the contrapositive. Let $\delta > 0$ and let G be a graph which cannot be made triangle-free by deleting at most $\delta|V(G)|^2$ edges. Define

$$\varepsilon := \delta/10$$

and

$$t = \lceil 1/\delta \rceil$$

and let

$$\gamma := \frac{(1-2\varepsilon)\varepsilon^3}{k_0(\varepsilon, t)^3}$$

where $k_0(\varepsilon, t)$ is as in Theorem 5.8. Note that, since both ε and t are defined in terms of δ, and $k_0(\varepsilon, t)$ is a function of ε and t, we have that γ is a function of δ. Let $X_1, \dots, X_k$ be the partition of $V(G)$ obtained by applying Theorem 5.8 to G with parameters ε and t. Let G' be the graph obtained from G by deleting

- every edge between X_i and X_j such that the pair (X_i, X_j) is not ε-regular,
- every edge between X_i and X_j such that $d(X_i, X_j) < 2\varepsilon$ and
- every edge fully contained within X_i for $1 \leq i \leq k$.

The number of edges deleted in the first step is at most

$$\varepsilon \binom{k}{2} \left\lceil \frac{|V(G)|}{k} \right\rceil^2 < \varepsilon |V(G)|^2.$$

Similarly, the number deleted in the second step is at most

$$2\varepsilon k^2 \left\lceil \frac{|V(G)|}{k} \right\rceil^2 < 4\varepsilon |V(G)|^2.$$

Finally, the number deleted in the third step is at most

$$\sum_{i=1}^{k} \binom{|X_i|}{2} \leq t \left(\frac{(|V(G)|/t)^2}{2} \right) \leq |V(G)|/2t.$$

Thus, the number of edges removed is at most

$$\left(5\varepsilon + \frac{1}{2t}\right) |V(G)|^2 \leq \delta |V(G)|^2.$$

By definition of G, the graph G' still contains a triangle, say xyz. By construction of G', we must have $x \in X_a$, $y \in X_b$ and $z \in X_c$ for some distinct indices a, b and c such that the pairs $(X_a, X_b), (X_a, X_c)$ and (X_b, X_c) are all ε-regular with density at least 2ε. However, by Lemma 5.7, the number of triangles with one vertex in each of the sets X_a, X_b and X_c is at least

$$(1 - 2\varepsilon)\varepsilon^3 |X_a||X_b||X_c| \geq \left(\frac{(1-2\varepsilon)\varepsilon^3}{k_0(\varepsilon, t)^3}\right) |V(G)|^3 = \gamma |V(G)|^3.$$

This completes the proof. □

The following corollary of the Triangle Removal Lemma is particularly useful; after proving it, we will illustrate it with a couple of applications. Basically, this says that if G is a graph with n vertices in which every edge is contained in at most one triangle, then the number of triangles is $o(n^2)$.

Corollary 5.11 (Ruzsa and Szemerédi [222])**.** *For every $\varepsilon > 0$ and $\ell \geq 1$ there exists $n_0 = n_0(\varepsilon, \ell)$ such that if G is a graph with $n \geq n_0$ vertices such that every edge of G is contained in at most ℓ triangles, then the number of triangles in G is at most εn^2.*

Proof. Let $\delta = \varepsilon/\ell$ and let $\gamma = \gamma(\delta)$ be as in Lemma 5.10. We set

$$n_0 := \left\lceil \frac{\ell}{6\gamma} \right\rceil.$$

Let G be any graph with $n \geq n_0$ vertices such that every edge of G is contained in at most ℓ triangles. Let $t(G)$ be the number of triangles in G

The proof has two steps; first, we prove a weaker upper bound on the number of triangles, and then we strengthen it. Firstly, for each edge e of G, let $t(e)$ be the number of triangles in G containing e. By assumption, we have $t(e) \leq \ell$ for every edge. The sum $\sum_{e \in E(G)} t(e)$ counts every triangle of G exactly three times. Therefore,

$$3t(G) = \sum_{e \in E(G)} t(e) \leq \ell |E(G)|$$

which implies that

$$t(G) \leq \frac{\ell}{3}\binom{n}{2} < \frac{\ell \cdot n^2}{6} = \frac{\ell \cdot n^3}{6n} \leq \frac{\ell \cdot n^3}{6n_0} \leq \gamma n^3.$$

Therefore, by the Triangle Removal Lemma, G can be made triangle-free by removing at most δn^2 edges. However, each edge e that is removed reduces the number of triangles by at most $t(e) \leq \ell$. Therefore, the total number of triangles which were contained in G to begin with must have been at most $\ell \delta n^2 = \varepsilon n^2$. □

As our first application of Corollary 5.11, let us derive the well-known $(6,3)$-Theorem. Recall that a *k-uniform hypergraph* is a pair (V, E) where V is a set of *vertices* and $E \subseteq \binom{V}{k}$ is a set of *hyperedges*. Note that a 2-uniform hypergraph is nothing more than a graph. The following can be seen as an upper bound for a Turán-type problem in 3-uniform hypergraphs where multiple different subhypergraphs are being forbidden.

Theorem 5.12 (The $(6,3)$-Theorem [222])**.** *For every $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon)$ such that if H is a 3-uniform hypergraph with $n \geq n_0$ vertices such that there does not a set $S \subseteq V(H)$ of cardinality six containing at least three hyperedges, then $|E(H)| \leq \varepsilon n^2$.*

Proof. Let n_0 be equal to $n_0(\varepsilon, 2)$ from Corollary 5.11. Let H be a 3-uniform hypergraph with $n \geq n_0$ vertices in which there is no set of six vertices containing at least three hyperedges. We let G be the graph with $V(G) = V(H)$ where uv is an edge of G if and only if there is a hyperedge in H containing u and v. Note that every hyperedge in H corresponds to a triangle in G, and so it suffices to show that G has at most εn^2 triangles.

To do this, we will show that no edge of G is contained in more than two triangles of G, from which the result will follow by Corollary 5.11. Let uv be any edge of G; our goal is to show that it is contained in at most two triangles. We break the proof into two cases; see Figure 5.2 for an illustration.

Case 1: There exist $e, e' \in E(H)$ such that $\{u, v\} \subseteq e$ and $|e \cap e'| = 2$.

We claim that there cannot exist a third hyperedge $e'' \in E(H) \setminus \{e, e'\}$ such that $e \cap e'' \neq \emptyset$. If such an e'' did exist, then $e \cup e' \cup e''$ would be a set of at most 6 vertices containing at least three distinct hyperedges of H (namely, e, e' and e'') which would contradict our assumption on H.

This implies that the four vertices of $e \cup e'$ have no neighbours outside of $e \cup e'$ and, moreover, that there is no edge in G from the unique vertex of $e \setminus e'$ to the unique vertex of $e' \setminus e$. Consequently, we get that the edge in $e \cap e'$ is contained in precisely two triangles and that all other edges in $e \cup e'$ are contained in only one triangle; in particular, uv is contained in at most two triangles.

Case 2: For every hyperedge e containing u and v and $e' \in E(H) \setminus \{e\}$, it holds that $|e \cap e'| \leq 1$.

Let e be any hyperedge containing u and v (which must exist, by definition of G). Suppose that there exists a vertex $x \notin e$ such that $\{u, v, x\}$ forms a triangle in G. Then there must exist hyperedges e' and e'' such that $u, x \in e'$ and $v, x \in e''$. By assumption, we must have $e \cap e' = \{u\}$ and $e \cap e'' = \{v\}$; in particular, e, e' and e'' are all distinct. The set $e \cup e' \cup e''$ is now a set of at most six vertices containing three distinct hyperedges, which is again a contradiction. This completes the proof. □

Another application of Corollary 5.11 is Roth's Theorem from combinatorial number theory. Essentially, this theorem says that any subset of $[n]$ without a non-trivial 3-term arithmetic progression has $o(n)$ elements; the following definition explains what we mean by this.

Definition 5.13. A sequence $(a_1, a_2, \ldots, a_k)$ of integers is called a *k-term arithmetic progression* if there exists $d \geq 0$ such that

$$a_{i+1} - a_i = d$$

for all $1 \leq i \leq k - 1$. It is *trivial* if $d = 0$ and *non-trivial* otherwise.

Theorem 5.14 (Roth's Theorem [220])**.** *For every $0 < \delta < 1$ there exists $N_0 = N_0(\delta)$ such that if $n \geq N_0$ and $A \subseteq [n]$ such that $|A| \geq \delta n$, then there exists a non-trivial 3-term arithmetic progression (a_1, a_2, a_3) such that $a_1, a_2, a_3 \in A$.*

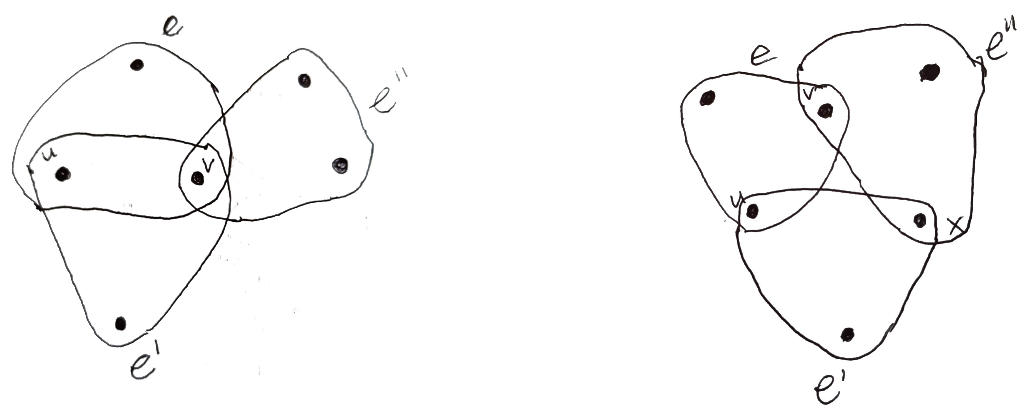


Figure 5.3: An illustration of the two cases in the proof of Theorem 5.12. Both of these configurations clearly consist of at most six vertices and at least three hyperedges.

Proof. For $0 < \delta < 1$, define
$$\varepsilon = \frac{\delta}{81}$$
and let $N_0 := n_0(\varepsilon, 1)/9$ where n_0 is as in Corollary 5.11. For $n \geq n_0$, we suppose that there exists $A \subseteq [n]$ of cardinality at least δn without any non-trivial 3-term arithmetic progression and derive a contradiction.

Let us start by building a graph G based on A as follows. Let $X = \{x_0, \ldots, x_{3n}\}$, $Y = \{y_0, \ldots, y_{3n}\}$ and $Z = \{z_0, \ldots, z_{3n}\}$ be disjoint sets of vertices.

- Add an edge from x_i to y_j if $j - i \in A$.
- Add an edge from y_i to z_j if $j - i \in A$.
- Add an edge from x_i to z_j if $(j - i)/2 \in A$.

(Note that, here, all arithmetic is viewed modulo $3n + 1$).

Let $t(G)$ denote the number of triangles in G. For $a \in A$ and $b \in \{0, \ldots, n\}$, we have that x_b, y_{b+a}, z_{b+2a} forms a triangle in G, by construction. Therefore,
$$t(G) \geq |A|(n+1) > \frac{\delta |V(G)|^2}{81} = \varepsilon |V(G)|^2.$$
(Here, we used that $|V(G)| = 9n + 3$ and $|A| \geq \delta n$).

Now, we will show that every edge of G is contained in at most one triangle. If this is true, then, by Corollary 5.11 and the fact that $|V(G)| \geq 9n \geq n_0(\varepsilon, 1)$, we will have that G contains at most $\varepsilon |V(G)|^2$ triangles, which will contradict the lower bound on the number of triangles proved above; thus, the proof will be complete.

Consider any triangle x_i, y_j, z_k in G with $x_i \in X, y_j \in Y$ and $z_k \in Z$. By construction, we must have
$$j - i \in A,$$

$$\frac{k-i}{2} \in A$$

and

$$k - j \in A.$$

Thus, $j-i$, $\frac{k-i}{2}$ and $k-j$ form a 3-term arithmetic progression in A with common difference $\frac{i-2j+k}{2}$. Since we are assuming that all 3-term arithmetic progressions in A are trivial, we must have

$$i - 2j + k = 0. \tag{5.15}$$

Thus, for any choice of two of i, j, k, the third is uniquely determined. Thus, every edge of G is contained in at most one triangle and we are done. □

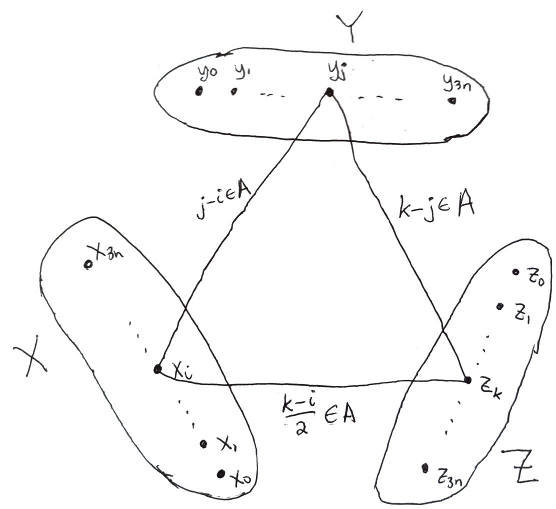


Figure 5.4: An illustration of the graph G constructed in the proof of Roth's Theorem.

5.3 Proof of the Regularity Lemma

Now that we have seen a few applications of the Regularity Lemma (Theorem 5.8), let's finally prove it. None of the individual steps of the proof are terribly difficult, but the whole proof is a bit long and technical. It is probably the most "involved" proof that we will see in this course. However, it is a powerful tool and, therefore, worth proving (and worth seeing the proof at least once in your life).

High-Level Overview. Before diving into details, let us take a moment to discuss the basic strategy of the proof. Let $\varepsilon > 0$ and $t \geq 1$ and define $t_0 := t$. Let G be any graph and let $n := |V(G)|$. Let $\mathcal{P}^0 = (X_1^0, \ldots, X_{t_0}^0)$ be an arbitrary partition of $V(G)$ into t_0 sets, each of cardinality $\lfloor n/t_0 \rfloor$ or $\lceil n/t_0 \rceil$.

Obviously, a silly arbitrary partition $\mathcal{P}^0$ will usually fail to do the job. In particular, it will fail to satisfy property (a) of Theorem 5.8. The first order of business will be to define a parameter

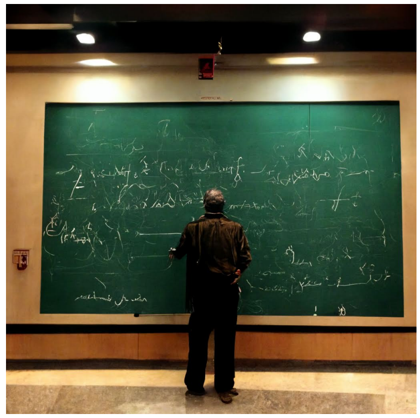


Figure 5.5: "Mathematician trying to prove the regularity lemma."

which "measures" how poor the partition $\mathcal{P}^0$ (or, any other partition) is at doing this job. The way that this will be measured is via the "mean square density" function f defined in Definition 5.16. The function f will have the property that $f(\mathcal{P})$ is between 0 and 1 for any partition $\mathcal{P}$. We think of a partition $\mathcal{P}'$ as being "better" than a partition $\mathcal{P}$ from the perspective of the mean-squared density function if $f(\mathcal{P}') > f(\mathcal{P})$.

The key to the proof is that, if many pairs of sets in the partition $\mathcal{P}^0$ fail to be ε-regular, then it is possible to refine $\mathcal{P}^0$ to obtain a partition $\mathcal{Q}^1$ of $V(G)$ into at most $t_0 2^{t_0-1}$ sets such that $f(\mathcal{Q}^1) \geq f(\mathcal{P}^0) + \varepsilon^5/2$. The sets in the partition $\mathcal{Q}^1$ may not be of close to the same cardinality as one another. However, by refining the partition further and doing a bit of "cleaning up," we will get a partition $\mathcal{P}^1$ with $f(\mathcal{P}^1) \geq f(\mathcal{P}^0) + \varepsilon^5/4$ such that the number t_1 of classes of $\mathcal{P}^1$ is at most $\left\lceil \frac{8t_0 2^{t_0-1}}{\varepsilon^5} \right\rceil$ and all the classes of $\mathcal{P}^1$ have size $\lfloor n/t_1 \rfloor$ or $\lceil n/t_1 \rceil$. If a large number of the pairs of sets in $\mathcal{P}^1$ still fail to be ε-regular, then we repeat this process, obtaining a partition $\mathcal{P}^2$ into $t_2 \leq \left\lceil \frac{8t_1 2^{t_1-1}}{\varepsilon^5} \right\rceil$ parts with mean square density at least $\varepsilon^5/4$ greater than that of $\mathcal{P}^1$, and so on. Since every partition of $V(G)$ has mean square density between zero and one, this procedure must terminate after at most $4/\varepsilon^5$ steps, yielding a partition which satisfies condition (a) of Theorem 5.8. Letting $m = \lfloor 4/\varepsilon^5 \rfloor$ be an upper bound on the number of iterations, the number of sets in the final partition is at most t_m where, for $1 \leq i \leq m$, the quantity t_i is defined by the recurrence $t_0 = t$ and $t_i := \left\lceil \frac{8t_{i-1} 2^{t_{i-1}-1}}{\varepsilon^5} \right\rceil$. Letting $k_0(\varepsilon, t) := t_m$, we see that property (c) of Theorem 5.8 holds as well.

Now, let's get into the details. Our first step is to formally define the mean square density of a partition.

Definition 5.16. Given a partition $\mathcal{P} = (X_1, \dots, X_t)$ of $V(G)$, define the *mean square density* of $\mathcal{P}$ to be

$$f(\mathcal{P}) := \frac{1}{n^2} \sum_{\substack{(i,j) \in [t]^2 \\ i \neq j}} |X_i||X_j| d(X_i, X_j)^2.$$

Remark 5.17. Let us try to explain why $f(\mathcal{P})$ is called the "mean square density." Let $x, y \in V(G)$

be two vertices chosen uniformly at random, independently of one another, and let

$$D := \begin{cases} d(X_i, X_j)^2 & \text{if } (x, y) \in X_i \times X_j \text{ for } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

Then it is easy to see that $f(\mathcal{P})$ is precisely the expected value (in other words, the mean) of D. So, in some sense, $f(\mathcal{P})$ computes the mean of the square of the density from X_i to X_j, where the pair (X_i, X_j) is chosen with probability proportional to $|X_i||X_j|$.

Observation 5.18. *Clearly, $f(A, B) \geq 0$ for any A and B. Therefore, $f(\mathcal{P}) \geq 0$ for any partition $\mathcal{P}$ of $V(G)$.*

A basic fact is that any partition has mean square density at most one.

Lemma 5.19. *For any partition $\mathcal{P}$ of $V(G)$,*

$$f(\mathcal{P}) \leq 1.$$

Proof. Exercise 5.13. □

Recall that, for partitions $\mathcal{P} = (X_1, \ldots, X_m)$ and $\mathcal{P}' = (X_1', \ldots, X_\ell')$ of $V(G)$, we say that $\mathcal{P}'$ *refines* $\mathcal{P}$ if for every $1 \leq i \leq \ell$ there exists $1 \leq j \leq m$ such that $X_i' \subseteq X_j$. Another important fact which we will use is that refining a partition does not decrease its mean square density. The following lemma will be used to compare the "contribution" from a pair (X_i, X_j) in a partition of $V(G)$ to the mean square density to the contribution of the subsets of X_i and X_j in a refinement of the partition.

Lemma 5.20. *If A and B are disjoint subsets of $V(G)$ and $(A_1, \ldots, A_s)$ and $(B_1, \ldots, B_t)$ are partitions of A and B, respectively, then*

$$|A||B|d(A, B)^2 \leq \sum_{i=1}^{s} \sum_{j=1}^{t} |A_i||B_j|d(A_i, B_j)^2.$$

Proof. By a simple double-counting argument, we see that

$$e(A, B) = \sum_{i=1}^{s} \sum_{j=1}^{t} e(A_i, B_j) = \sum_{i=1}^{s} \sum_{j=1}^{t} |A_i||B_i|d(A_i, B_j). \tag{5.21}$$

Squaring both sides yields

$$e(A, B)^2 = \left(\sum_{i=1}^{s} \sum_{j=1}^{t} |A_i||B_i|d(A_i, B_j) \right)^2.$$

For each $(i, j) \in [s] \times [t]$, let

$$x_{i,j} := \sqrt{|A_i||B_j|}$$

and

$$y_{i,j} := d(A_i, B_j)\sqrt{|A_i||B_j|}.$$

By Hölder's Inequality (Lemma C.5), we get that

$$\left(\sum_{i=1}^{s}\sum_{j=1}^{t}|A_i||B_i|d(A_i,B_j)\right)^2 = \left(\sum_{i=1}^{s}\sum_{j=1}^{t}|x_{i,j}y_{i,j}|\right)^2 \leq \left(\sum_{i=1}^{s}\sum_{j=1}^{t}x_{i,j}^2\right)\left(\sum_{i=1}^{s}\sum_{j=1}^{t}y_{i,j}^2\right)$$

$$= \left(\sum_{i=1}^{s}\sum_{j=1}^{t}|A_i||B_j|\right)\left(\sum_{i=1}^{s}\sum_{j=1}^{t}|A_i||B_j|d(A_i,B_j)^2\right)$$

It is easily observed that

$$\sum_{i=1}^{s}\sum_{j=1}^{t}|A_i||B_j| = |A||B|. \tag{5.22}$$

Therefore,

$$e(A,B)^2 \leq |A||B|\left(\sum_{i=1}^{s}\sum_{j=1}^{t}|A_i||B_j|d(A_i,B_j)^2\right).$$

Dividing both sides by $|A||B|$ and recalling that $d(A,B) = \frac{e(A,B)}{|A||B|}$ completes the proof. □

Lemma 5.23. *If $\mathcal{P}'$ and $\mathcal{P}$ are partitions of $V(G)$ such that $\mathcal{P}'$ refines $\mathcal{P}$, then*

$$f(\mathcal{P}) \leq f(\mathcal{P}').$$

Proof. Write $\mathcal{P} = (X_1,\ldots,X_m)$ and $\mathcal{P}' = (X_1',\ldots,X_\ell')$. For $1 \leq i \leq m$, let

$$R_i := \{j : 1 \leq r \leq \ell \text{ and } X_r' \subseteq X_i\}.$$

Then, by Lemma 5.20 to every pair (X_i,X_j) with $i \neq j$, we get

$$f(\mathcal{P}) = \frac{1}{n^2}\sum_{i\neq j}|X_i||X_j|d(X_i,X_j)^2 \leq \frac{1}{n^2}\sum_{i\neq j}\sum_{r\in R_i}\sum_{s\in R_j}|X_r'||X_s'|d(X_r',X_s')^2 = f(\mathcal{P}')$$

as desired. □

Note that the inequality

$$\left(\sum_{i=1}^{n}|x_iy_i|\right)^2 \leq \left(\sum_{i=1}^{n}x_i^2\right)\left(\sum_{i=1}^{n}y_i^2\right)$$

in Lemma C.5 is tight if and only if the vectors $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$ are constant multiples of one another. What does this tell us in the context of the proof of Lemma 5.20? It tells us that the "gain" in the mean square density will be zero if and only if there exists d such that $d(A_i,B_j) = d$ for all $(i,j) \in [s]\times[t]$. Of course, if such a d exists, then it must be the case that $d = d(A,B)$. Thus, if (A,B) is not ε-regular and we refine A and B in a way that "pulls out" some subsets of A and B of density different from $d(A,B)$, then we should be able to get a "boost" in mean square density. The following lemma quantifies the size of this "boost."

Lemma 5.24. *Suppose A and B are disjoint subsets of $V(G)$ and let $A_1 \subseteq A$ and $B_1 \subseteq B$. Define $A_2 := A \setminus A_1$ and $B_2 := B \setminus B_1$. Then*

$$\sum_{1\le i,j\le 2} |A_i||A_j|d(A_i,A_j)^2 \ge |A||B|d(A,B)^2 + (d(A_1,B_1)-d(A,B))^2|A_1||B_1|.$$

Proof. We have

$$|A_1||B_1|(d(A_1,B_1)-d(A,B))^2 \le \sum_{1\le i,j\le 2} |A_i||B_j|(d(A_i,B_j)-d(A,B))^2.$$

Expanding $(d(A_i,B_j)-d(A,B))^2$ in the sum on the right side (for each $1 \le i,j \le 2$ individually) yields the following expression:

$$\sum_{1\le i,j\le 2} |A_i||B_j|d(A_i,B_j)^2 - 2d(A,B)\sum_{1\le i,j\le 2} |A_i||B_j|d(A_i,B_j) + d(A,B)^2 \sum_{1\le i,j\le 2} |A_i||B_j|.$$

Applying (5.21) to the second summation and (5.22) to the third, we get that this is equal to

$$\sum_{1\le i,j\le 2} |A_i||B_j|d(A_i,B_j)^2 - |A||B|d(A,B)^2.$$

We now add $|A||B|d(A,B)^2$ to both sides to get the desired bound. □

Okay, let's describe an algorithm for producing the partition in Theorem 5.8. Recall that $t_0 := t$ and $\mathcal{P}^0$ is an arbitrary partition of $V(G)$ into sets $X_1^0, \ldots, X_{t_0}^0$, each of size roughly n/t_0. If $\mathcal{P}^0$ already satisfies condition (a) of Theorem 5.8, then we have nothing more to do, and we simply output it.

So, assume that condition (a) of Theorem 5.8 is not satisfied. For each pair (X_i^0, X_j^0) which is not ε-regular and such that $i \ne j$, take $X_{i,j} \subseteq X_i^0$ and $X_{j,i} \subseteq X_j^0$ such that $|X_{i,j}| \ge \varepsilon|X_i^0|$, $|X_{j,i}| \ge \varepsilon|X_j^0|$ and $|d(X_{i,j}, X_{j,i}) - d(X_i^0, X_j^0)| \ge \varepsilon$. Let $X'_{i,j} := X_i^0 \setminus X_{i,j}$ and $X'_{j,i} := X_j^0 \setminus X_{j,i}$. By Lemma 5.24, if we replace X_i^0 and X_j^0 with four parts $X_{i,j}, X'_{i,j}, X_{j,i}, X'_{j,i}$, then the contribution of the pair (X_i^0, X_j^0) to the mean square density increases by at least[2]

$$\frac{(d(X_{i,j},X_{j,i}) - d(X_i^0,X_j^0))^2|X_{i,j}||X_{j,i}|}{n^2} \ge \frac{\varepsilon^2\cdot\varepsilon|X_i^0|\varepsilon|X_j^0|}{n^2} \ge \frac{\varepsilon^4\lfloor n/t_0\rfloor^2}{n^2} > \frac{\varepsilon^4}{2t_0^2}.$$

Also, crucially, by Lemma 5.20, any partitions of X_i^0 and X_j^0 that refine the partitions $(X_{i,j}, X'_{i,j})$ and $(X_{j,i}, X'_{j,i})$ will also give us an extra contribution of at least $\frac{\varepsilon^4}{2t_0^2}$ to the mean square density compared with the contribution from (X_i^0, X_j^0). So, if there are more than εt_0^2 non-ε-regular pairs and we refine every non-ε-regular pair in the way described above (at the same time), then we get a total boost of at least

$$\left(\frac{\varepsilon^4}{2t_0^2}\right)(\varepsilon t_0^2) \ge \frac{\varepsilon^5}{2}$$

to the mean square density.

Now, let $\mathcal{Q}^1$ be the partition obtained from the refinement procedure described in the previous paragraph. The following lemma explains how to refine $\mathcal{Q}^1$ further and rearrange it so that all sets have roughly the same size without decreasing the mean square density by very much.

[2]Note that the n^2 on the denominator comes from the original definition of the mean square density in Definition 5.16.

Lemma 5.25. *Let $\delta > 0$ and $\mathcal{Q} = (X_1, \ldots, X_m)$ be a partition of $V(G)$. Then there exists a partition $\mathcal{P} = (X'_1, \ldots, X'_t)$ of $V(G)$ such that*

- $t = \lceil \frac{2m}{\delta} \rceil$,
- $||X'_i| - |X'_j|| \leq 1$ *for all* $1 \leq i, j \leq t$ *and*
- $f(\mathcal{P}) \geq f(\mathcal{Q}) - \delta$.

Proof. Let $t = \lceil \frac{2m}{\delta} \rceil$ and let $0 \leq r \leq t-1$ such that

$$|V(G)| = t \left\lfloor \frac{n}{t} \right\rfloor + r.$$

Let $\tilde{\mathcal{Q}}$ be an arbitrary refinement of $\mathcal{Q}$ obtained by breaking up each class X_i of $\mathcal{Q}$ into sets of size $\lfloor n/t \rfloor$ or $\lceil n/t \rceil$ and one "leftover" set $Y_i \subseteq X_i$ with $|Y_i| < n/t$. When doing so, one needs to be careful to maintain the restriction that the total number of sets of size $\lfloor n/t \rfloor$ is at most $t - r$ and the number of size $\lceil n/t \rceil$ is at most r. By Lemma 5.20,

$$f(\tilde{\mathcal{Q}}) \geq f(\mathcal{Q}).$$

Now, we let $\mathcal{P}$ be a partition obtained from $\tilde{\mathcal{Q}}$ by putting all of the sets of leftover vertices together and breaking them up into sets of size $\lfloor n/t \rfloor$ or $\lceil n/t \rceil$ abitrarily. Note that it is possible to do this since our partitioning of the rest of the vertices uses at most $t - r$ sets of size $\lfloor n/t \rfloor$ and at most r of size $\lceil n/t \rceil$.

The last step is to show that the mean square density of $\mathcal{P}$ cannot be much smaller than that of $\tilde{\mathcal{Q}}$. For each $1 \leq i \leq m$, the number of leftover vertices in the partition of X_i is at most n/t. Since $d(A, B) \leq 1$ for any two disjoint non-empty sets A and B, the contribution of these leftover sets to the mean square density of $\tilde{\mathcal{Q}}$ is at most

$$2 \sum_{i=1}^{m} \frac{|Y_i| n}{n^2} \leq \frac{2}{n} \sum_{i=1}^{m} |Y_i| \leq 2m \left(\frac{n}{t} \right) \leq \delta.$$

Thus, when we put together all these leftover vertices and divide them up differently, the most that we can lose from the mean square density is δ. This completes the proof of the lemma. □

So, the partition $\mathcal{P}^1$ is obtained from $\tilde{\mathcal{Q}}^1$ by applying Lemma 5.25 with $\delta = \frac{\varepsilon^5}{4}$. To finish the proof, all that we need to notice is that the procedure for obtaining $\mathcal{P}^1$ from $\mathcal{P}^0$ that we have just described can be applied in exactly the same way to obtain $\mathcal{P}^2$ from $\mathcal{P}^1$, and so on, until we finally reach a partition which satisfies property (a) of Theorem 5.8. Since every iteration increases the mean square density by at least $\varepsilon^5/4$, this will terminate after at most $4/\varepsilon^5$ steps. The total number of sets in the final partition will be bounded above by t_m where $m = \lfloor 4/\varepsilon^5 \rfloor$, $t_0 = t$ and $t_i = \left\lceil \frac{8t_{i-1} 2^{kt_{i-1}-1}}{\varepsilon^5} \right\rceil$ for all $i \geq 1$.

Remark 5.26. Notice, here, that the number of sets in the final partition is, in the worst case, completely enormous. In particular, we have

$$k_0(\varepsilon, t) = t_m \geq 2^{t_{m-1}} \geq 2^{2^{t_{m-2}}} \geq 2^{2^{2^{t_{m-3}}}} \geq 2^{2^{2^{2^{\cdot^{\cdot^{\cdot^{2^{t_0}}}}}}}}.$$

That is, the number of sets in the final partition may be as large as a "tower of 2s" of height $\varepsilon^{-O(1)}$.

5.4 Exercises

5.1 Let $0 < \varepsilon < 1$, let G be a graph. Suppose that (X_1, X_2) is a partition of $V(G)$ such that the pair (X_1, X_2) is ε-regular in G. Show that (X_1, X_2) is ε-regular in the complement $\overline{G}$ of G.

5.2 Prove that if G is a graph with n vertices and at most $\varepsilon^3 \left\lfloor \frac{n^2}{4} \right\rfloor$ edges, then G has an ε-regular partition with exactly two parts.

5.3 Let G be a graph and (X_1, X_2, X_3) be a partition of $V(G)$. Suppose that

$$0 < 2\varepsilon < \min\{d(X_i, X_j), 1 - d(X_i, X_j)\}$$

for all distinct $i, j \in \{1, 2, 3\}$ and that all of the pairs (X_i, X_j) are ε-regular. Let P_3 be the path on three vertices and, for $i \in \{1, 2, 3\}$, let v_i be the ith vertex on the path. Show that the number of induced copies of P_3 in which v_i is in X_i for $1 \le i \le 3$ is at least

$$(1 - 2\varepsilon)(d(X_1, X_2) - \varepsilon)(1 - d(X_1, X_3) - \varepsilon)(d(X_2, X_3) - \varepsilon)|X_1||X_2||X_3|.$$

Hint: Apply Exercise 5.1 and Theorem 5.7.

5.4 State and prove a version of Lemma 5.7 for counting copies of K_4, where $V(G)$ is partitioned into four sets X_1, X_2, X_3 and X_4.

5.5 Consider the following version of the Triangle Removal Lemma for deleting vertices instead of edges: *For every $0 < \delta < 1$ there exists $0 < \gamma < 1/6$ such that if G is a graph with n vertices and at most γn^3 triangles, then G can be made triangle-free by deleting at most δn vertices.* Disprove this statement.

5.6 The goal of this exercise is to disprove a natural generalization of the Counting Lemma (Lemma 5.7) for 3-uniform hypergraphs.[3] Let n be divisible by 4 and let $V = V_1 \sqcup V_2 \sqcup V_3 \sqcup V_4$ where $|V_i| = n/4$ for $1 \le i \le 4$.

(a) Let H be a 3-uniform hypergraph with $V(H) = V$ such that, for distinct i, j, k, if $(x, y, z) \in V_i \times V_j \times V_k$, then $E(H)$ contains $\{x, y, z\}$ with probability 1/8 independently of all other triples. Compute the expected number of sets $S \subseteq V(G)$ such that $|S| = 4$ and $\binom{S}{3} \subseteq E(H)$.

(b) Let G be a graph with vertex set $V(G) = V$ where, if $u \in V_i$ and $v \in V_j$ with $i \ne j$, then uv is in $E(G)$ with probability 1/2 independently of any other pair. Let H' be a hypergraph with vertex set V where $\{x, y, z\} \in E(H')$ if x, y and z form a triangle in G. For distinct i, j, k and $(x, y, z) \in V_i \times V_j \times V_k$, determine the probability that $\{x, y, z\} \in E(H')$.

(c) Let H' be the hypergraph as in the previous part of the question. Compute the expected number of sets $S \subseteq V(G)$ such that $|S| = 4$ and $\binom{S}{3} \subseteq E(H')$.

(d) Write a natural generalization of Lemma 5.7 and explain why the hypergraphs H and H' disprove that statement. In your explanation, you can make some reasonable probabilistic assumptions without proving them.

[3]Note that there is a different generalization of the counting lemma for hypergraphs that is true.

5.7 Consider the following extension of the Triangle Removal Lemma to r-uniform hypergraphs:

Hypergraph Removal Lemma: *Let $r \geq 2$. For every $0 < \delta < \frac{1}{r!}$ there exists $\gamma = \gamma(\delta)$ with $0 < \gamma < \frac{1}{(r+1)!}$ such that if H is an r-uniform hypergraph with at most γn^{r+1} copies of $K_{r+1}^{(r)}$, then H can be made $K_{r+1}^{(r)}$-free by deleting at most δn^r hyperedges.*

(a) Use the Hypergraph Removal Lemma to prove the following extension of the Ruzsa–Szemerédi Theorem:

Let $r \geq 2$. For every $\varepsilon > 0$ and $\ell \geq 1$ there exists $n_0 = n_0(r, \varepsilon, \ell)$ such that if H is an r-uniform hypergraph such that every hyperedge of H is contained in at most ℓ copies of $K_{r+1}^{(r)}$, then the number of copies of $K_{r+1}^{(r)}$ in G is at most εn^r.

(b) Use the extension of the Ruzsa–Szemerédi Theorem from part (a) to prove the following:

Szemerédi's Theorem: *For every $k \geq 3$ and $\delta > 0$ there exists $N_0(k, \delta)$ such that if $n \geq N_0(k, \delta)$, then every set $A \subseteq [n]$ with $|A| \geq \delta n$ contains a non-trivial k-term arithmetic progression.*

(c) Use the extension of the Ruzsa–Szemerédi Theorem from part (a) to prove the following:

Corners Theorem: *For every $\delta > 0$ there exists $N_0(\delta)$ such that if $n \geq N_0(\delta)$, then, for every $A \subseteq [n] \times [n]$ with $|A| \geq \delta n^2$, there exist $x, y, z \in [n]$ such that*

$$\{(x, y), (x, y + z), (x + z, y)\} \subseteq A.$$

5.8 Prove that, for every $\varepsilon > 0$, there exists $n_0 = n_0(\varepsilon)$ such that if G is a graph on $n \geq n_0$ vertices in which every edge is contained in exactly one triangle, then $|E(G)| \leq \varepsilon n^2$.

5.9 (a) Let G be a graph. A set $M \subseteq E(G)$ is called an *matching* if the edges in M are pairwise disjoint. We say that M is an *induced matching* if it is a matching and, additionally, for any distinct $e_1, e_2 \in M$ there are no edges of G linking an endpoint of e_1 to an endpoint of e_2. Prove the following:

For every $\varepsilon > 0$ there exists $n_0 = n_0(\varepsilon)$ such that if G is a bipartite graph with $n \geq n_0$ vertices and $E(G) = M_1 \cup \cdots M_n$ where the sets $M_1, \ldots, M_n$ are disjoint induced matchings in G, then $|E(G)| \leq \varepsilon n^2$.

Hint: Use Corollary 5.11 with $t = 1$.

(b) Let G be a graph with n vertices, where n is divisible by 3, such that every edge of G is contained in a unique triangle. Using a probabilistic argument, or otherwise, show that there is a bipartite graph G' with $2n/3$ vertices and at least $\frac{2|E(G)|}{27}$ edges such that the edges of G' can be partitioned into $n/3$ disjoint matchings.

5.10 Let n and m be integers such that $3\binom{m}{2} + m < n$. Show that if $A \subseteq [n]$ such that $|A| = m$ and A does not contain a non-trivial 3-term arithmetic progression, then there exits $x \in [n] \setminus A$ such that $A \cup \{x\}$ does not contain a non-trivial 3-term arithmetic progression.

5.11 For every $1 \leq k \leq n$, find a simple formula for the number of non-trivial k-term arithmetic progressions in $[n]$ (i.e. not in terms of a summation).

5.12 Using the Regularity Lemma, prove that the number of n-vertex triangle-free graphs is $2^{(1/4+o(1))n^2}$.

5.13 Prove Lemma 5.19.

5.14 Using the Regularity Lemma (Theorem 5.8), prove that one can instead obtain a partition into k parts $X_1, \dots, X_k$ (the value of k might be different) such that for every $1 \le i \le k$ there are at most εk indices $j \ne i$ such that (X_i, X_j) is not ε-regular.

Hint: Apply the Regularity Lemma for some $\varepsilon' < \varepsilon$ of your choosing to get a partitioning $X'_1, \dots, X'_{k'}$. How many of the parts X'_i can be involved in more than $\varepsilon k'$ pairs which are not ε-regular? What happens if we incorporate the vertices of these parts into the other parts?

5.15 Prove that, for every $\varepsilon > 0$ and $t \in \mathbb{N}$, there exists $k_0(\varepsilon, t)$ such that, for every graph G, there is a partition $X_0, X_1, \dots, X_k$ of $V(G)$ and a spanning[4] subgraph G' of G such that

- $|X_0| \le \varepsilon |V(G)|$,
- $|X_1| = |X_2| = \cdots = |X_k|$,
- G' contains no edges within X_i for any $0 \le i \le k$,
- the density between X_i and X_j in G' is either 0 or at least 10ε,
- (X_i, X_j) is ε-regular in G' for all $1 \le i \ne j \le k$, and
- each vertex $v \in V(G)$ satisfies $d_{G'}(v) \ge d_G(v) - 1000\varepsilon|V(G)|$.

Hint: Apply Exercise 5.14.

5.16 Let H be a graph. An *H-tiling* of a graph G is a collection $\mathcal{H}$ of disjoint subsets of $V(G)$ such that every set in $\mathcal{H}$ contains a copy of H. Say that an H-tiling $\mathcal{H}$ *covers* a vertex $v \in V(G)$ if there exists a set in $\mathcal{H}$ containing v. If $\mathcal{H}$ does not cover v, say that it *misses* v.

(a) Let $K_{a,b,c}$ be the complete tripartite graph with parts of sizes a, b and c. Assume that $a \le b \le c$. Show that, for all $\gamma > 0$, there exists a graph G such that $\delta(G) \ge \left(\frac{2a+b+c}{2(a+b+c)} - \gamma\right)|V(G)|$ and every $K_{a,b,c}$-tiling of G misses at least $\gamma\left(\frac{1}{b+c} + \frac{1}{a}\right)|V(G)|$ vertices of G.

(b) Given a graph H, let $\sigma(H)$ be the minimum of $\left|f^{-1}(\{1\})\right|$ over all proper colourings $f : V(H) \to \{1, \dots, \chi(H)\}$ of H. Generalize the construction in part **??** to find a functions $\phi(H)$ and $\tau(H)$, depending on $\sigma(H)$, $\chi(H)$ and $|V(H)|$, such that for all $\gamma > 0$, there exists a graph G such that $\delta(G) \ge (\phi(H) - \gamma)|V(G)|$ such that every H-tiling of G misses at least $\gamma\tau(H)|V(G)|$ vertices of G. Make $\phi(H)$ as large as you can.

5.17 Prove that, for every $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that if G is an n-vertex bipartite graph with bipartition (A, B) with $||A| - |B|| < \varepsilon n$, the pair (A, B) is ε-regular and $d(A, B) > 10\varepsilon$, then G contains a matching M of size at least $\left(\frac{1}{2} - \delta\right)n$.

Hint: Hall's Theorem for matchings in bipartite graphs. If you don't know it, look it up!

5.18 Our goal is to prove the following theorem of Thomassen [260] and show that it is asymptotically tight:

Thomassen's Theorem. *For every $\delta > 0$ there exists $C = C(\delta)$ such that, if G is a triangle-free graph with n vertices and minimum degree at least $\left(\frac{1}{3} + \delta\right)n$, then $\chi(G) \le C$.*

[4] A *spanning subgraph* of G is a subgraph G' of G such that $V(G') = V(G)$.

Let $0 < \delta < 1/2$ and let G be a graph satisfying the hypotheses of Thomassen's Theorem. Define $\varepsilon := \frac{\delta}{1000}$, $t = \lceil 1000/\delta \rceil$ and let $X_1, \ldots, X_k$ be a partition of $V(G)$ obtained from applying the Regularity Lemma to G, ε and t. Let $d = 50\varepsilon$ and, for $I \subseteq [k]$, let

$$X_I := \{v \in V(G) : |N(v) \cap X_i| \geq d|X_i| \iff i \in I\}.$$

Note that the sets X_I for $I \subseteq [k]$ partition $V(G)$. In all of the following questions, you may assume that $\delta n > 10^{100}k$.

(a) Prove that, if $|I| \leq 2k/3$, then X_I is an independent set.

(b) Prove that, if $|I| \geq 2k/3$, then $X_I = \emptyset$.

(c) Deduce Thomassen's Theorem.

(d) The goal in this question is to show that Thomassen's Theorem is asymptotically tight. Prove that there exists a sequence $G_1, G_2, G_3, \ldots$ of graphs such that

- G_n is triangle-free for all $n \geq 1$,
- $\lim_{n\to\infty} \frac{\delta(G_n)}{|V(G_n)|} = \frac{1}{3}$ where $\delta(G_n)$ is the minimum degree of G_n, and
- $\lim_{n\to\infty} \chi(G_n) = \infty$.

Hint: Look up the Mycielski construction and apply it. You may need to solve some recurrence relations.

5.5 Challenge Problems

5.1* Prove that, for every $\delta > 0$, there exist $n_0 = n_0(\delta)$ such that, for $n \geq n_0$, any set $A \subseteq [n]^2$ with at least δn^2 elements contains a triple of the form $(x, y), (x + d, y), (x, y + d)$ for $d > 0$. Also, show that this implies Roth's Theorem.

5.2* Recall that $\alpha(G)$ is the cardinality of the largest independent set in G (i.e. the largest set of vertices containing no edges). Show that if G is a K_4-free graph with n vertices satisfying $\alpha(G) = o(n)$, then

$$|E(G)| \leq (1/8 + o(1))\, n^2.$$

Chapter 6

Independent Sets, Matchings and Trees

The purpose of this chapter is to touch on a few beautiful techniques for counting structures in graphs. Several of the proofs are based on the notion of "entropy," which is a concept from information theory. We will also provide a basic application of a powerful modern technique known as the container method.

6.1 Independent Sets in Triangle-Free Graphs

The first question that we will focus on is: given a triangle-free graph G on n vertices, how large of an independent set must G contain? Let's start by obtaining a relatively easy bound of roughly $\sqrt{n}$. Recall that $\alpha(G)$ is the cardinality of the largest independent set in a graph G.

Proposition 6.1. *If G is a triangle-free graph with n vertices, then*

$$\alpha(G) \geq \sqrt{n} - \frac{1}{2}.$$

Proof. Since G is triangle-free, for every vertex v, the neighbourhood $N(v)$ is an independent set. Therefore, if G contains a vertex of degree at least $\sqrt{n} - \frac{1}{2}$, then we are done.

So, we assume that the maximum degree of G is less than $\sqrt{n} - \frac{1}{2}$. Let $f : V(G) \to \mathbb{N}$ be a function obtained by ordering the vertices of G by $v_1, \ldots, v_n$ arbitrarily and letting $f(v_i)$ be the smallest integer that is not contained in $\{f(v_j) : j < i \text{ and } v_j \in N(v_i)\}$. Then, by construction, the set $f^{-1}(c)$ is independent for every $c \in \mathbb{N}$. Also, the neighbours of any given vertex v are coloured with at most $d(v)$ distinct colours, and so $f(v) \leq d(v) + 1 \leq \sqrt{n} + \frac{1}{2}$. for every vertex v. Therefore, by the Pigeonhole Principle, there exists $1 \leq c \leq \sqrt{n} + \frac{1}{2}$ such that

$$|f^{-1}(c)| \geq \frac{n}{\sqrt{n} + \frac{1}{2}} > \sqrt{n} - \frac{1}{2}$$

which completes the proof. □

Remark 6.2. The bound in the above proof can be improved a wee bit if we use "Brooks' Theorem" (which is not covered in this course) instead of the simple greedy colouring procedure. However, the gain would not be very large.

The argument in Proposition 6.1 was not terribly sophisticated, but it is somewhat instructive. In particular, it tells us that a simple way of finding a large independent set in a triangle-free graph is just by taking the neighbourhood of a large degree vertex. Next, we present an argument which is based on the same fundamental principle, but contains more clever analysis.

Theorem 6.3 (Shearer's Bound [236])**.** *If G is a triangle-free graph with n vertices, then*

$$\alpha(G) \geq (1 + o(1))\sqrt{\frac{n\ln(n)}{2}}$$

as $n \to \infty$.

Proof. Let $f : [0, \infty) \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x = 0, \\ 1/2 & \text{if } x = 1, \\ \frac{x\ln(x) - x + 1}{(x-1)^2} & \text{otherwise.} \end{cases}$$

We start by making some observations about f. By taking a couple of derivatives, it is not hard to that f satisfies

$$1 - f(x)(x + 1) + f'(x)(x - x^2) = 0 \tag{6.4}$$

and that f is convex (see Appendix C). Therefore, it satisfies

$$f(x) \geq f(y) + f'(y)(x - y) \tag{6.5}$$

for every $x, y \in [0, 1]$.

We show that every triangle-free graph G on n vertices has an independent set of cardinality at least $f(2|E(G)/n)n$. We will do this by induction on n, where the case $n = 1$ is trivial.

Define

$$d = 2|E(G)|/n = \sum_{v \in V(G)} d(v)/n.$$

That is, d is the *average degree* of G. What we are trying to prove is that $\alpha(G) \geq f(d)n$. For each vertex v, let

$$d_2(v) := \sum_{u \in N(v)} d(u)$$

and let G_v be the graph obtained from G by deleting v and all of its neighbours. Note that the number of vertices in G_v is $n - d(v) - 1$. Define

$$d_v := 2|E(G_v)|/(n - d(v) - 1).$$

That is, d_v is the average degree of G_v.

Since G is triangle-free, we have that, for every $v \in V(G)$,

$$|E(G)| = |E(G_v)| + d_2(v) \tag{6.6}$$

Also, for any independent set S in G_v, the set $S \cup \{v\}$ is independent in G. Therefore, $\alpha(G) \geq \alpha(G_v) + 1$. Applying this bound for every $v \in V(G)$, induction on $|V(G)|$ and (6.5), we get

$$n \cdot \alpha(G) \geq \sum_{v \in V(G)} (\alpha(G_v) + 1) \geq n + \sum_{v \in V(G)} f(d_v)(n - d(v) - 1)$$

$$\geq n + \sum_{v \in V(G)} [f(d) + f'(d)(d_v - d)](n - d(v) - 1).$$

Doing a bit of rearranging on the right side (and using the fact that $\sum_{v\in V(G)} d(v) = dn$) yields

$$n + \sum_{v \in V(G)} [f(d) + f'(d)(d_v - d)](n - d(v) - 1)$$

$$= n + \sum_{v \in V(G)} [f(d) - f'(d)d](n - d(v) - 1) + \sum_{v \in V(G)} f'(d)d_v(n - d(v) - 1)$$
$$= n + (f(d) - f'(d)d)(n^2 - dn - n) + f'(d) \sum_{v \in V(G)} d_v(n - d - 1).$$

By definition, we have $d_v(n - d - 1) = 2|E(G_v)|$ for all $v \in V(G)$. Combining this with (6.6) and the fact that $|E(G)| = nd$, we get that the above expression is equal to

$$n + (f(d) - f'(d)d)(n^2 - dn - n) + f'(d) \sum_{v \in V(G)} (nd - 2d_2(v))$$

$$= n + f(d)(n^2 - dn - n) + f'(d)(d^2n + dn) - 2f'(d) \sum_{v \in V(G)} d_2(v).$$

Now, $\sum_{v\in V(G)} d_2(v)$ precisely counts the number of ways to choose a vertex v of G, choose a neighbour u of v and then choose a neighbour w of u. This is the same as the number of ways to simply choose the vertex u first and choose two of its neighbours, v and w. Thus,

$$\sum_{v \in V(G)} d_2(v) = \sum_{u \in V(G)} d(u)^2 \geq \frac{1}{n} \left(\sum_{u \in V(G)} d(u) \right)^2 = nd^2$$

by Corollary C.7. So, putting everything together (and dividing by n) we get that

$$\alpha(G) \geq f(d)n + 1 - f(d)(d + 1) + f'(d)(d - d^2)$$
$$= f(d)n$$

by (6.4).

Recall that G has an independent set of cardinality $d(v)$ for any $v \in V(G)$. So, it has an independent set of cardinality at least the average degree d. So, by the result above,

$$\alpha(G) \geq \max\{d, f(d)n\}.$$

If $d \geq \sqrt{\frac{n\ln(n)}{2}}$, then we are done. Otherwise, $d < \sqrt{\frac{n\ln(n)}{2}}$, and plugging this into $f(d)n$ yields a lower bound of roughly

$$\approx \left(\frac{\ln(d)}{d}\right) n \geq \left(\frac{\ln\left(\sqrt{\frac{n\ln(n)}{2}}\right)}{\sqrt{\frac{n\ln(n)}{2}}} \right) n \geq \sqrt{\frac{n\ln(n)}{2}}$$

(since $\ln\left(\sqrt{\frac{n\ln(n)}{2}}\right) \geq \frac{1}{2}\ln(n)$). $\square$

For $k, \ell \in \mathbb{N}$, the *Ramsey number* $R(k,\ell)$ is the minimum N such that, in any colouring of the edges of K_N with red and blue, there is either a red copy of K_k or a blue copy of K_ℓ. In this language, Theorem 6.3 can be thought of as saying that

$$R(3,k) \leq (1+o(1))\frac{k^2}{\log(k)}$$

as $k \to \infty$. The topic of Ramsey numbers (including a proof of their existence in general) is covered in more depth in Math 422 at UVic [68, Chapter 6].

6.2 Counting Independent Sets in Regular Graphs

Figure 6.1: "A container with a fingerprint."

Recall that a graph G is *d-regular* if every vertex of G has degree d. In this section, we focus on the problem of determining how many independent sets a d-regular graph can have.

Example 6.7. Suppose that n is a multiple of $2d$ and consider the graph $H_{d,n}$ obtained by taking a disjoint union of $\frac{n}{2d}$ copies of $K_{d,d}$. The number of independent sets in $K_{d,d}$ is

$$2^{d+1} - 1$$

(pick one side of the bipartition and then pick a non-empty subset of that side; subtract one because the set $\emptyset$ was counted twice). Therefore, the number of independent sets in $H_{d,n}$ is

$$\left(2^{d+1} - 1\right)^{\frac{n}{2d}}.$$

For large d, this is roughly

$$2^{(1/2+O(1/d))n}.$$

Our goal will be to prove an upper bound on the number of independent sets which is somewhat close (but not quite as good as) the value achieved in Example 6.7. In proving this result, we will use that any set of more than $n/2$ vertices in a d-regular graph must contain a reasonably large number of edges.

Lemma 6.8. *If G is a d-regular graph and $S \subseteq V(G)$ has cardinality at least $n/2+t$, then there are at least td edges within S.*

Proof. Exercise 6.4 (c). $\square$

Theorem 6.9 (Alon [4])**.** *For every $\varepsilon > 0$ there exists $d_0 = d_0(\varepsilon)$ such that if $d \geq d_0$, then every d-regular graph on n vertices has at most*

$$2^{(1/2+\varepsilon)n}$$

independent sets.

Proof (Sapozhenko [225]). Fix $0 < \varepsilon < 1$. Let $m = m(\varepsilon)$ be an integer chosen large enough so that

$$m \cdot e < 2^{\varepsilon m/3} \tag{6.10}$$

and then let $d = d(\varepsilon)$ be large enough so that

$$d > \frac{3m}{2\varepsilon}. \tag{6.11}$$

Our goal is to show that every d-regular graph on n vertices has at most $2^{(1/2+\varepsilon)n}$ independent sets.

Let G be any such graph. Fix an arbitrary ordering $v_1, \dots, v_n$ on the vertices of G. For each independent set S of G, we construct a set F_S called the *fingerprint* of S and a set C_S called the *container* of S by applying the following algorithm:

1. Initialize $C_S := V(G)$ and $F_S = \emptyset$.
2. While there exists a vertex in C_S with at least m neighbours in C_S, let v_i be a vertex of C_S such that $|N(v_i) \cap C_S|$ is maximum and, subject to this, the index i is minimum.
   a) If $v_i \notin S$, then update $C_S := C_S \setminus \{v_i\}$ and go back to 2.
   b) If $v_i \in S$, then update $F_S := F_S \cup \{v_i\}$ and $C_S := C_S \setminus (\{v\} \cup N(v))$ and go back to 2.

When this algorithm terminates, every vertex of C_S has at most m neighbours in C_S. Therefore, the number of edges within C_S is at most $|C_S| \cdot m/2 \leq nm/2$. Thus, by Lemma 6.8,

$$|C_S| \leq \frac{n}{2} + \frac{n \cdot m}{2d}. \tag{6.12}$$

Each time a vertex is added to F_S, the cardinality of C_S decreases by more than m. Therefore,

$$|F_S| \leq \frac{n}{m}. \tag{6.13}$$

Also, by construction, for every independent set S,

$$F_S \subseteq S \subseteq F_S \cup C_S. \tag{6.14}$$

This is because every element of S that was deleted from C_S was added to F_S.

The crucial, but extremely subtle, observation is that any two independent sets S and T such that $F_S = F_T$ must also satisfy $C_S = C_T$. Indeed, if you think through the algorithm, and apply induction, you can see that, if $F_S = F_T$, then the algorithm will select the same vertex in each step and behave in exactly the same way with respect to S or T. So, we can let g be a function mapping subsets of $V(G)$ of cardinality at most n/m to sets of cardinality at most $n/2 + nm/(2d)$ with the property that

$$g(F_S) = C_S \quad \text{for every independent } S \subseteq V(G).$$

For each set F of cardinality at most n/m, let $\mathcal{I}_F$ be the collection of independent sets S of G such that $F_S = F$. Then, by (6.12), (6.13) and (6.14), the number of independent sets in G is at most

$$\begin{aligned}
\sum_{F:|F|\leq n/m} |\mathcal{I}_F| &\leq \sum_{F:|F|\leq n/m} |\{S \subseteq V(G) : F \subseteq S \subseteq g(F)\}| \\
&= \sum_{F:|F|\leq n/m} \left[\sum_{j=0}^{|g(F)|-|F|} \binom{|g(F)| - |F|}{j} \right] \\
&\leq \sum_{i=0}^{n/m} \left[\binom{n}{i} \sum_{j=0}^{n/2-i} \binom{n/2 + nm/2d - i}{j} \right].
\end{aligned}$$

Recall that, for any $n \geq 1$ (see Proposition 6.15 below for a proof),

(1) $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.

Therefore, for any i, the "inner summation" can be bounded above as follows

$$\sum_{j=0}^{n/2-i} \binom{n/2 + nm/2d - i}{j} \leq 2^{n/2+nm/2d}.$$

The following inequalities are also standard (again, see Proposition 6.15 for proofs).

(2) $\binom{n}{k-1} \leq \binom{n}{k}$ for $1 \leq k \leq n/2$ and

(3) $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$.

Therefore,

$$\sum_{i=0}^{n/m} \binom{n}{i} \leq (n/m + 1) \left(\frac{ne}{n/m} \right)^{n/m} = (n/m + 1)\,(me)^{n/m}$$

and so the number of independent sets in G is at most

$$(n/m + 1)\,(me)^{n/m}\, 2^{n/2+nm/2d}.$$

Finally, by (6.10) and (6.11), this is at most

$$2^{(1/2+\varepsilon)n}$$

where, here, we also use the fact that $n > d > m$ and so, by (6.10), we also have $n/m+1 < 2^{\varepsilon n/3}$ $\square$

For the sake of completeness, let's quickly prove the three simple inequalities involving binomial coefficients that we needed in the proof of Theorem 6.9.

Proposition 6.15. *The statements (1)–(3) are true.*

Proof. The quantity 2^n counts the number of subsets of $[n]$ and $\binom{n}{k}$ counts the number of subsets of $[n]$ of cardinality exactly k. Therefore,

$$2^n = \sum_{i=0}^{n} \binom{n}{i}$$

and so (1) holds.

For every subset S of $[n]$ of cardinality $k-1$, there are $n-k+1$ elements of $[n]$ that can be added to S to obtain a subset of $[n]$ of size k. Likewise, for every subset of cardinality k, there are k elements that can be removed to get a subset of cardinality $k-1$. Therefore, by double counting,

$$(n-k+1)\binom{n}{k-1} = k\binom{n}{k}. \tag{6.16}$$

So, (2) holds.

Finally, for (3), iterating the bound in (6.16) and using the fact that $\binom{n}{0} = 1$ gives us

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

Therefore,

$$\binom{n}{k} = \prod_{i=0}^{k-1} \frac{n-i}{k-i} \leq \frac{n^k}{k!}.$$

Now, recall that $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ for any $x \in \mathbb{R}$. Thus,

$$e^k = \sum_{n=0}^{\infty} \frac{k^n}{n!} \geq \frac{k^k}{k!}.$$

Putting these bounds together completes the proof. □

Recall that a *proper q-colouring* of a graph G is a mapping $f : V(G) \to \{1, \ldots, q\}$ such that if $uv \in E(G)$, then $f(u) \neq f(v)$. In Exercise 6.8, we will follow an argument similar to the proof of Theorem 6.9 to prove the following bound on the number of q-colourings of a d-regular graph.

Theorem 6.17 (Galvin [116])**.** *Let $q \geq 2$ be fixed. For every $\varepsilon > 0$ there exists $d_0 = d_0(\varepsilon)$ such that if $d \geq d_0$, then every d-regular graph on n vertices has at most*

$$q^n \left(\frac{1}{2} + \varepsilon\right)^n$$

proper q-colourings.

Figure 6.2: “Information.”

6.3 An Introduction to Entropy

The remaining proofs in this chapter will be more probabilistic in nature than any of the other proofs given in this course so far. However, it is all still relatively basic probability theory, requiring not much more than an understanding of expectation and conditional probability. Specifically, they will use the notion of the “entropy” of a random variable, which is defined as follows.

Definition 6.18. Let X be a discrete random variable with finite range S.[1] The *entropy* of X is defined to be

$$H(X) = \sum_{x \in S} \mathbb{P}(X = x) \log_2 \left(\frac{1}{\mathbb{P}(X = x)} \right).$$

Remark 6.19. Occasionally, the base of the logarithm will be important; however, most of the time, it is not.

A nice way to think of the quantity $\log\left(\frac{1}{\mathbb{P}(X=x)}\right)$ is as quantifying the amount of “surprise” that you feel from seeing that $X = x$. If $X = x$ is something that happens with probability close to 1, i.e. it is very likely, then the amount of surprise that you feel from seeing that $X = x$ is close to zero. E.g., the fact that the sun comes up in the morning and sets in the evening is never really surprising because the probability that it would not happen is so low. Accordingly, if $\mathbb{P}(X = x)$ is close to 1, then $\log\left(\frac{1}{\mathbb{P}(X=x)}\right)$ is close to zero. Conversely, it is surprising when unlikely events happen, and this is reflected by the fact that $\log\left(\frac{1}{\mathbb{P}(X=x)}\right)$ is large when $\mathbb{P}(X = x)$ is small. Given this intuition, one can think of entropy $H(X)$ as being the “expected surprise” of the random variable X.

Example 6.20. Suppose that X is uniformly distributed on a finite set S. That is, for every $x \in S$, we have $\mathbb{P}(X = x) = \frac{1}{|S|}$. Then

$$H(X) = \sum_{x \in S} \mathbb{P}(X = x) \log \left(\frac{1}{\mathbb{P}(X = x)} \right) = \sum_{x \in S} \frac{\log(|S|)}{|S|} = \log(|S|).$$

[1] Here, the *range* of X is understood to be the set of x such that $\mathbb{P}(X = x) \neq 0$.

So, in general, if X is uniformly distributed on a finite set, then $H(X)$ is the logarithm of the cardinality of the range of X.

We will need to build up some basic facts about entropy. Many of these are related to the notion of "conditional" probabilities, random variables and entropy; we start by giving the relevant definitions. Anyone who has taken Math 352, or any other probability or stats course, at UVic will be familiar with several of these definitions.

Definition 6.21. Given two events A and B in a probability space such that $\mathbb{P}(B) \neq 0$, the *conditional probability* of A *given* B is defined by

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Definition 6.22. Say that events A and B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Observation 6.23. *If $\mathbb{P}(B) \neq 0$, then A and B are independent if and only if $\mathbb{P}(A \mid B) = \mathbb{P}(A)$.*

Definition 6.24. Two discrete random variables X and Y on the same probability space are *independent* if the events $\{X = x\}$ and $\{Y = y\}$ are independent for every x and y.

Definition 6.25. Given discrete random variables X and Y and y in the range of Y, let the *conditional entropy of X given $Y = y$* be given by

$$H(X \mid Y = y) = \sum_{x} \mathbb{P}(X = x \mid Y = y) \log\left(\frac{1}{\mathbb{P}(X = x \mid Y = y)}\right)$$

where this sum is over all x such that $\mathbb{P}(X = x \mid Y = y) \neq 0$.

Definition 6.26. For discrete random variables X and Y, the *conditional entropy of X given Y* is

$$H(X \mid Y) = \sum_{y} \mathbb{P}(Y = y) H(X \mid Y = y).$$

With all of these new definitions, its time for a few examples.

Example 6.27. Suppose that X and Y are independent. Then, for any y in the range of Y, we have

$$\mathbb{P}(X = x \mid Y = y) = \mathbb{P}(X = x).$$

Therefore, $H(X \mid Y = y)$ is simply equal to $H(X)$. Also,

$$H(X \mid Y) = \sum_{y} \mathbb{P}(Y = y) H(X \mid Y = y) = \sum_{y} \mathbb{P}(Y = y) H(X) = H(X).$$

The entropy $H(X)$ of X can be thought of as the amount of "information" carried by the random variable X. With this interpretation in mind, what Example 6.27 is illustrating is that, if X and Y are independent, then, by "revealing" the information that $Y = y$, we do not reduce the amount of information contained in X. That is, if X and Y are independent, then knowing that $Y = y$ tells us nothing about the value of X. We are still equally "surprised" by seeing the value of X as we would be if we had known nothing about Y. An example to think about is if X is the indicator function of you winning the lottery and Y is the indicator function of you forgetting to take out the garbage. Whether or not you forgot to take out the garbage has no effect on how surprised you would be to win the lottery, as these two events have nothing to do with one another.

The next example illustrates another extreme. Say that a random variable X is *determined* by a random variable Y if, for every y in the range of Y, there exists $f(y)$ such that $\mathbb{P}(X = f(y) \mid Y = y) = 1$.

Example 6.28. Suppose that X is determined by Y. Then, for each y in the range of Y, we have that

$$\begin{aligned} H(X \mid Y = y) &= \sum_x \mathbb{P}(X = x \mid Y = y) \log\left(\frac{1}{\mathbb{P}(X = x \mid Y = y)}\right) \\ &= 1 \log(1) = 0. \end{aligned}$$

Therefore, we also have that $H(X \mid Y) = 0$.

Example 6.28 illustrates the fact that, when X is determined by Y, if we already know the value of Y, then there is no additional surprise from learning the value of X. Thus, $H(X \mid Y) = 0$.

Example 6.29. Let Y be uniformly distributed on $\{0, 1\}$. Let X be the random variable such that, if $Y = 0$, then X is uniformly distributed on $\{4, 5\}$ and, if $Y = 1$, then X is uniformly distributed on $\{6, 7\}$. If we think about X in isolation, then we see that it is equally likely to be equal to any of the values in $\{4, 5, 6, 7\}$. Therefore, by the result of Example 6.20,

$$H(X) = \log(4).$$

Also, by definition,

$$H(Y) = \log(2).$$

Now, let's think about $H(X \mid Y)$. By going through a calculation similar to Example 6.20, we see that $H(X \mid Y = 0) = H(X \mid Y = 1) = \log(2)$. Thus, we have

$$H(X \mid Y) = \log(2).$$

Note that, if we think about $H(Y \mid X)$ instead, then the situation is different. Knowing the value of X completely determines Y; therefore, $H(Y \mid X) = 0$.

The following lemma collates several of the basic facts that we will need about entropy and conditional entropy.

Lemma 6.30. *Let $X, Y, Z, X_1, \ldots, X_n$ be discrete random variables and let S be the range of X. The following hold:*

(a) $H(X) \leq \log(|S|)$ where equality holds if and only if X is uniformly distributed on S,

(b) $H(X_1, \ldots, X_n) = H(X_1) + H(X_2 \mid X_1) + \cdots + H(X_n \mid X_1, \ldots, X_{n-1})$,

(c) if Z is determined by Y, then $H(Y, Z) = H(Y)$ and

(d) $H(X \mid Y) \leq H(X)$, with equality if and only if X and Y are independent.

Proof. We have already seen in Example 6.20 that the uniform distribution has entropy equal to the log of the size of its range. The inequality in (a) follows from Jensen's Inequality (Theorem C.4) and the fact that the function $x\log(x)$ is convex. The case of equality follows from the fact that $x\log(x)$ is actually strictly convex; we omit the details.

Now, let's think about property (b). First, consider the case $n = 2$. In this case,

$$H(X_1, X_2) = \sum_{(x_1,x_2)} \mathbb{P}((X_1, X_2) = (x_1, x_2)) \log\left(\frac{1}{\mathbb{P}((X_1, X_2) = (x_1, x_2))}\right)$$

$$= \sum_{(x_1,x_2)} \mathbb{P}(X_1 = x_1 \text{ and } X_2 = x_2) \log\left(\frac{1}{\mathbb{P}(X_1 = x_1 \text{ and } X_2 = x_2)}\right)$$

$$= \sum_{(x_1,x_2)} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2 \mid X_1 = x_1) \log\left(\frac{1}{\mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2 \mid X_1 = x_1)}\right)$$

$$= \sum_{(x_1,x_2)} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2 \mid X_1 = x_1) \left[\log\left(\frac{1}{\mathbb{P}(X_1 = x_1)}\right) + \log\left(\frac{1}{\mathbb{P}(X_2 = x_2 \mid X_1 = x_1)}\right)\right].$$

Now, split this into two separate sums. The first one is

$$\sum_{x_1}\sum_{x_2} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2 \mid X_1 = x_1) \log\left(\frac{1}{\mathbb{P}(X_1 = x_1)}\right) = H(X_1)$$

and the second is

$$\sum_{x_1}\sum_{x_2} \mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2 \mid X_1 = x_1) \log\left(\frac{1}{\mathbb{P}(X_2 = x_2 \mid X_1 = x_1)}\right) = H(X_2 \mid X_1)$$

as desired. Now, for $n \geq 3$, we proceed by induction. Let $Y_1 = (X_1, X_2)$ and, for $2 \leq i \leq n-1$, let $Y_i = X_{i+1}$. Then, by induction,

$$H(Y_1, \dots, Y_{n-1}) = H(Y_1) + H(Y_2 \mid Y_1) + \cdots + H(Y_{n-1} \mid Y_1, \dots, Y_{n-2}).$$

But we also know that $H(Y_1) = H(X_1, X_2) = H(X_1) + H(X_2 \mid X_1)$. So, if we substitute this for $H(Y_1)$ and write out the above expression in terms of $X_1, \dots, X_n$, then we get the result.

For part (c), see Exercise 6.12. For part (d), see Exercise 6.13. □

By combining Lemma 6.30 (b) and (d), we get that, for any random variables $X_1, \dots, X_n$,

$$H(X_1, \dots, X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i \mid X_1, \dots, X_{i-1}) \leq \sum_{i=1}^{n} H(X_i).$$

The following result, known as Shearer's Lemma, can be seen as a generalization of this simple fact. We will use it to prove two nice theorems shortly.

Lemma 6.31 (Shearer's Lemma [44])**.** *Let $t \geq 1$ and let $\mathcal{F} \subseteq 2^{[n]}$ such that every $i \in [n]$ is contained in at least t sets in $\mathcal{F}$. Then, for any random variables $X_1, \dots, X_n$,*

$$H(X_1, \dots, X_n) \leq \frac{1}{t} \sum_{F \in \mathcal{F}} H(X_i : i \in F).$$

Proof. Let F be an arbitrary element of $\mathcal{F}$ and let the elements of F be labelled as $v_1, \dots, v_k$ where $v_1 < v_2 < \dots < v_k$. Then, by Lemma 6.30 (b),

$$H(X_i : i \in F) = H(X_{v_1}, \dots, X_{v_k}) = \sum_{i=1}^{k} H(X_{v_i} \mid X_{v_1}, \dots, X_{v_{i-1}}).$$

Now, by Lemma 6.30 (d), we have that, for any $1 \le i \le k$,

$$H(X_{v_i} \mid X_{v_1}, \dots, X_{v_{i-1}}) \ge H(X_{v_i} \mid X_1, \dots, X_{v_i - 1}).$$

So, if we lower bound each term of the sum of $H(X_i : i \in F)$ over all $F \in \mathcal{F}$ in this way, then the term $H(X_v \mid X_1, \dots, X_{v-1})$ appears at least t times for any $v \in [n]$. Therefore,

$$\sum_{F \in \mathcal{F}} H(X_i : i \in F) \ge t \sum_{v=1}^{k} H(X_v \mid X_1, \dots, X_{v-1}) = tH(X_1, \dots, X_n)$$

as desired. □

The following is a simple consequence of Shearer's Lemma, known as Han's Inequality.

Corollary 6.32 (Han's Inequality [128])**.** *For any random variables $X_1, \dots, X_n$,*

$$H(X_1, \dots, X_n) \le \frac{1}{n-1} \sum_{j=1}^{n} H(X_i : i \ne j).$$

Proof. Apply Shearer's Lemma to the collection $\mathcal{F} := \{[n] \setminus \{j\} : 1 \le j \le n\}$. Clearly, each $i \in [n]$ is contained in exactly $n - 1$ of the sets in $\mathcal{F}$. □

As a few quick application of Han's Inequality, let us prove the well-known Loomis–Whitney Inequality and an edge-isoperimetric inequality in the hypercube.

Theorem 6.33 (The Loomis–Whitney Inequality [181])**.** *Let B be a measurable body in $\mathbb{R}^n$ and, for $1 \le j \le n$, let π_j be the projection operator onto the plane $\{(x_1, \dots, x_n) : x_j = 0\}$. Denote the measure of B by $\mu(B)$ and the measure of $\pi_j(B)$, viewed as a subset of $\mathbb{R}^{n-1}$, by $\mu'(\pi_j(B))$. Then*

$$\mu(B) \le \prod_{j=1}^{n} \mu'(\pi_j(B))^{1/(n-1)}.$$

Proof. For each integer $k \ge 1$, let

$$\mathcal{C}_k := \left\{ \prod_{j=1}^{n} \left[\frac{\ell_j}{k}, \frac{\ell_j + 1}{k} \right) : \ell_1, \dots, \ell_n \in \mathbb{Z} \right\}.$$

That is, it is the set of all "half open" cubes with side-lengths $1/k$ whose bottom corner is in the set $\{\ell/k : \ell \in \mathbb{Z}\}^n$. Let B_k be the union of all unit cubes $C \in \mathcal{C}_k$ such that $C \cap B_k \ne \emptyset$. Then

$$\mu(B) = \lim_{k \to \infty} \mu(B_k). \tag{6.34}$$

A similar phenomenon holds for all of the projections of B as well. The rest of the proof will focus on B_k as opposed to B; at the end, we will use (6.34) to translate the result back to B.

Let $B_k = \bigsqcup_{i=1}^N C_i$ where $C_i \in \mathcal{C}_k$ for all $1 \leq i \leq k$. Each cube C_i has measure k^{-n} and so $\mu(B_k) = N \cdot k^{-n}$. Choose $i \in [N]$ uniformly at random and let $X = (X_1, \ldots, X_n)$ where X_j is the jth coordinate of the bottom corner of the cube C_i. Then X is a uniform random variable whose range has cardinality $N = k^n \mu(B_k)$. Thus, by Lemma 6.30 (a),

$$\log(k^n \mu(B_k)) = H(X).$$

By Han's Inequality,

$$H(X) \leq \frac{1}{n-1} \sum_{j=1}^{n} H(X_i : i \neq j).$$

Now, for fixed $1 \leq j \leq n$, we have that $H(X_i : i \neq j)$ is a random variable whose range has cardinality precisely equal to $k^{n-1}\mu'(\pi_j(B_k))$. Applying Lemma 6.30 (a) again, we get

$$H(X_i : i \neq j) \leq \log(\mu'(\pi_j(B_k))).$$

Putting this all together, we get that

$$k^n \mu(B_k) \leq \prod_{j=1}^{n} \left(k^{n-1}\mu'(\pi_j(B_k))\right)^{1/(n-1)}$$

$$\Rightarrow \mu(B_k) \leq \prod_{j=1}^{n} \mu'(\pi_j(B_k))^{1/(n-1)}$$

The result follows by taking the limit as $k \to \infty$ and applying (6.34) and the analogous statements for the projections of B. $\square$

Example 6.35. If $B = \prod_{i=1}^n [0, a_i]$ is an axis-aligned cube, then $\mu(B) = \prod_{i=1}^n a_i$ and $\mu'(B_j) = \prod_{i \neq j} a_i$ for all $1 \leq j \leq n$. Therefore, every such B satisfies the bound in Theorem 6.33 with equality.

Theorem 6.36 (Edge-Isoperimetry Inequality in the Hypercube)**.** *Let Q_d be the d-dimensional hypercube and let $A \subseteq \{0,1\}^d$ be non-empty. Then*

$$e(A) \leq \frac{|A|}{2} \log_2(|A|).$$

Proof. Let $X = (X_1, \ldots, X_d)$ be a uniformly random vertex of A where, for $1 \leq i \leq d$, X_i is the ith coordinate of X. By Lemma 6.30 (a) and Han's Inequality, we have

$$\log_2(|A|) = H(X) = nH(X) - (n-1)H(X)$$

$$\geq nH(X) - \sum_{j=1}^{n} H(X_i : i \neq j) = \sum_{j=1}^{n} [H(X) - H(X_i : i \neq j)].$$

By Lemma 6.30 (b), this is equal to

$$\sum_{j=1}^{n} H(X_j \mid X_i : i \neq j).$$

For each $1 \leq j \leq n$, let S_j be the range of $(X_i : i \neq j)$. In other words, S_j is the set of all sequences of the form $(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)$ where $(x_1, \ldots, x_n) \in A$. Going all the way back to the definition of conditional entropy, we get that

$$H(X_j \mid X_i : i \neq j) = \sum_{(x_i : i \neq j) \in S_j} \mathbb{P}((X_i : i \neq j) = (x_i : i \neq j)) H(X_j \mid (X_i : i \neq j) = (x_i : i \neq j)).$$

Given $x = (x_1, \ldots, x_n) \in \{0,1\}^d$, let x^j be the vertex of the hypercube obtained from x by changing the jth coordinate from 0 to 1 or vice versa. Then, for any such pair $\{x, x^j\} \subseteq \{0,1\}^d$, we have

$$\mathbb{P}((X_i : i \neq j) = (x_i : i \neq j)) = \frac{|\{x \cap x^j\} \cap A|}{|A|}.$$

Also, the number of choices for X_j given that $(X_i : i \neq j) = (x_i : i \neq j)$ is equal to $|\{x \cap x^j\} \cap A|$. Thus, assuming that $(x_i : i \neq j) \in S_j$, we have

$$H(X_j \mid (X_i : i \neq j) = (x_i : i \neq j)) = \log_2(|\{x \cap x^j\} \cap A|)$$

which is equal to $\log_2(1) = 0$ if $|\{x \cap x^j\} \cap A| = 1$ and $\log_2(2) = 1$ otherwise. Thus, $H(X_j \mid X_i : i \neq j)$ is equal to $\frac{2}{|A|}$ times the number of pairs $\{x, x^j\} \subseteq \{0,1\}^d$ such that both x and x^j belong to A. In other words, $H(X_j \mid X_i : i \neq j)$ is precisely equal to the number of edges within A whose endpoints differ on the jth coordinate. Therefore,

$$\sum_{j=1}^{n} H(X_j \mid X_i : i \neq j) = \frac{2e(A)}{|A|}.$$

Putting this all together gives us

$$\log(|A|) \geq \frac{2e(A)}{|A|}$$

and we are done by solving for $e(A)$. $\square$

Example 6.37. If A induces a hypercube of dimension k within Q_d, then $|A| = 2^k$ and

$$e(A) = k2^{k-1} = \frac{|A|}{2} \log_2(|A|)$$

and so A satisfies the bound in Theorem 6.36 with equality.

6.4 Permanents and Counting Perfect Matchings

Recall that a *perfect matching* in a graph G is a set $M \subseteq E(G)$ of edges such that every vertex is contained in exactly one edge of M. The question that we address here is: how many perfect matchings can a graph with a given degree sequence have? Our first example is similar to the key example from Section 6.2.

Example 6.38. Let $d_1, \ldots, d_k$ be positive integers and let G be a disjoint union of complete bipartite graphs K_{d_i,d_i} for $1 \leq i \leq k$. The number of perfect matchings in K_{d_i,d_i} is precisely $d_i!$ and so the number in G is

$$\prod_{i=1}^{k} d_i!.$$

Let (A, B) be the bipartition of G. Then, in the ith component of G, there are d_i vertices of A, each of degree $d(v)$. Another way to express this number is as follows. Then the number of perfect matchings in G is precisely

$$\prod_{v \in A} (d(v)!)^{1/d(v)}$$

or

$$\prod_{v \in V(G)} (d(v)!)^{1/(2d(v))}.$$

These may seem like weird ways of writing things, but this will be convenient for comparing this example to the bounds that we will prove below.

Given a graph G, let $\mathrm{pm}(G)$ be number of perfect matchings in G. Our goal will be to show that the bound in Example 6.38 is actually tight. The special case of bipartite G is particularly interesting. The proof combines some of the facts that we have learned about entropy in the previous section.

Theorem 6.39 (Brègman's Theorem [36])**.** *If G is a bipartite graph with bipartition (A, B), then*

$$\mathrm{pm}(G) \leq \prod_{v \in A} (d(v)!)^{1/d(v)}.$$

Proof (Radhakrishnan [212]). Let G be a bipartite graph with bipartition (A, B). Let M be a perfect matching in G chosen uniformly at random from the set of all perfect matchings. Then, by the result of Example 6.20, we have that the entropy of M is equal to the logarithm of the cardinality of its range. Therefore,

$$H(M) = \log(\mathrm{pm}(G)).$$

So, proving an upper bound on $\mathrm{pm}(G)$ is equivalent to proving an upper bound on $H(M)$; we focus on the latter.

For each $v \in A$, let X_v be equal to the vertex $u \in B$ such that $vu \in M$. Note that the joint random variable

$$X = (X_v : v \in A)$$

is equivalent to M; that is, X is determined by M and M is determined by X as well. Therefore, $H(X) = H(M) = \log(\mathrm{pm}(G))$ and we can focus on bounding the entropy of X. For each $u \in B$, we can similarly define Y_u to be the vertex of A such that $Y_u u \in M$.

For any total order $\prec$ on A, we have, by Lemma 6.30 (b),

$$H(X) = \sum_{v \in A} H(X_v \mid X_w, w \prec v). \tag{6.40}$$

Given a vertex $v \in A$ and a total order $\prec$ on A, we define the following random variable

$$R_v^{\prec} := \left| N(v) \setminus \{u \in B : Y_u \prec v\} \right|.$$

Note that $R_v^{\prec}$ is determined by $(X_w : w \prec v)$. So, by Lemma 6.30 (c) and (d) and the definition of conditional entropy,

$$H(X_v \mid X_w : w \prec v) = H(X_v \mid (X_w : w \prec v), R_v^{\prec}) \leq H(X_v \mid R_v^{\prec}) = \sum_{k=1}^{d(v)} \mathbb{P}(R_v^{\prec} = k) H(X_v \mid R_v^{\prec} = k).$$

When $R_v^{\prec} = k$, then the number of choices for X_v is at most k. So, by Lemma 6.30 (a),

$$H(X_v \mid X_w : w \prec v) \leq \sum_{k=1}^{d(v)} \mathbb{P}(R_v^{\prec} = k)\log(k).$$

Plugging this back into equation (6.40) and summing over all choices of $\preceq$ yields

$$H(X) \leq \frac{1}{n!}\sum_{\prec}\sum_{v\in A}\sum_{k=1}^{d(v)} \mathbb{P}(R_v^{\prec} = k)\log(k) = \frac{1}{n!}\sum_{v\in A}\sum_{k=1}^{d(v)}\log(k)\sum_{\prec}\mathbb{P}(R_v^{\prec} = k)$$

$$= \frac{1}{n!}\sum_{v\in A}\sum_{k=1}^{d(v)}\log(k)\sum_{\prec}\sum_{M}\frac{1_{\{R_v^{\prec}=k\}}}{\mathrm{pm}(G)}$$

$$= \frac{1}{\mathrm{pm}(G)}\sum_{v\in A}\sum_{k=1}^{d(v)}\log(k)\sum_{M}\sum_{\prec}\frac{1_{\{R_v^{\prec}=k\}}}{n!}$$

$$= \frac{1}{\mathrm{pm}(G)}\sum_{v\in A}\sum_{k=1}^{d(v)}\log(k)\sum_{M}\frac{1}{d(v)}$$

$$= \sum_{v\in A}\sum_{k=1}^{d(v)}\log(k)\frac{1}{d(v)}$$

$$= \sum_{v\in A}\frac{\log(d(v)!)}{d(v)}$$

Therefore,

$$\mathrm{pm}(G) \leq \prod_{v\in A}(d(v)!)^{1/d(v)}.$$

□

Brègman's Theorem can also be interpreted as an upper bound on the permanent of a matrix. Given an $n \times n$ matrix $A = (a_{i,j})_{1\leq i,j\leq n}$, the *permanent* of A is

$$\mathrm{per}(A) := \sum_{\sigma\in S_n}\prod_{i=1}^{n} a_{i,\sigma(i)}$$

where S_n is the set of all bijections $\sigma : \{1,\ldots,n\} \to \{1,\ldots,n\}$ (i.e. the *symmetric group* of order n). See also [68, Chapter 3]. Note that this is the same thing as the determinant, except without alternating signs. For completeness, define $\mathrm{per}(A) = 0$ if A is not a square matrix. Given a bipartite graph G with bipartition (A, B) with $A = \{u_1,\ldots,u_k\}$ and $B = \{v_1,\ldots,v_\ell\}$, the *biadjacency matrix* of G is the $k \times \ell$ matrix where the (i, j) entry is 1 if $u_iv_j \in E(G)$ and 0 otherwise. The following is not hard to prove (see Exercise 7 in Chapter 3 of the course notes for Math 422 at UVic [68]).

Observation 6.41. *If G is a bipartite graph with biadjacency matrix A, then*

$$\mathrm{pm}(G) = \mathrm{per}(A).$$

Thus, Theorem 6.39 is equivalent to the following.

Corollary 6.42 (Brègman's Theorem, for permanents [36]). *If A is a $\{0,1\}$-valued matrix with row sums $r_1, \ldots, r_n$, then*

$$\mathrm{per}(A) \leq \prod_{i=1}^{n} (r_i!)^{1/r_i}.$$

What about non-bipartite graphs, you may ask? As it turns out, there is a nice trick for applying Corollary 6.42 to get the following best-possible upper bound for general graphs; we will explore the proof in the exercises.

Theorem 6.43 (Kahn and Lovász, unpublished). *For any graph G,*

$$\mathrm{pm}(G) \leq \prod_{v \in V(G)} (d(v)!)^{1/2d(v)}.$$

Proof. Exercise 6.16. □

6.5 Homomorphism Density of Trees

Figure 6.3: "Tree."

Recall that a graph T is called a *tree* if it is connected and does not contain a cycle. It is easily observed that T is a tree if and only if there is a unique path between any two vertices of T, and that every tree satisfies $|E(T)| = |V(T)| - 1$. In this section, we will apply the facts that we have learned about entropy to prove the following result which is akin to Theorem 4.8, but with the graph C_4 replaced by a tree.[2] Recall the definition of $t(H, G)$ from Chapter 4.

Theorem 6.44 (Sidorenko [238]). *For every tree T and graph G,*

$$t(T, G) \geq t(K_2, G)^{|E(T)|}.$$

[2]Note that this theorem solves Challenge Problem 4.1*.

The proof will involve constructing a random homomorphism from T into a graph G and comparing its entropy to that of a random homomorphism from K_2 to G. Thus, it will be useful to understand some basic properties of random homomorphisms from K_2 to G.

Claim 6.45. *Let G be a graph with at least one edge and let g be a uniformly random homomorphism from K_2 to G. Then, for any $x \in V(K_2)$ and $v \in V(G)$,*

$$\mathbb{P}(g(x) = v) = \frac{d(v)}{2|E(G)|}.$$

Also, for any $xy \in E(K_2)$ and $uv \in E(G)$,

$$\mathbb{P}(g(y) = v \mid g(x) = u) = \frac{1}{d(u)}.$$

Proof. The number of homomorphisms from K_2 to G is $2|E(G)|$. The number of such homomorphisms that satisfy $g(x) = v$ is $d(v)$. So, the probability that a random homomorphism satisfies $g(x) = v$ is $\frac{d(v)}{2|E(G)|}$.

For the second part of the claim, again, note that the number of homomorphisms g with $g(x) = u$ is $d(u)$. Since g is uniform, any of these homomorphisms should be equally likely; thus, the probability that $g(y) = v$ given that $g(x) = u$ is $\frac{1}{d(u)}$. □

Next, we show that, if the vertices of a tree T are mapped into $V(G)$ in a way that follows a "branching random walk," then the resulting homomorphism will have properties similar to those described in Claim 6.45.

Proposition 6.46. *For any tree T and graph G with at least one edge, there is a distribution on homomorphisms from T to G such that if f is chosen according to that distribution, then, for any $x \in V(T)$ and $v \in V(G)$,*

$$\mathbb{P}(f(x) = v) = \frac{d(v)}{2|E(G)|}.$$

Also, for any $xy \in E(T)$ and $uv \in E(G)$,

$$\mathbb{P}(f(y) = v \mid f(x) = u) = \frac{1}{d(u)}.$$

Proof. Let r be an arbitrary *root* vertex of T. For each $x \in V(T)\setminus\{r\}$, the *parent* of x is the unique neighbour $p(x)$ of x that is closer to r. We say that x is a *child* of y if $p(x) = y$. We construct f as follows. Firstly, for the root, we let $f(r)$ be a random vertex of G chosen according to the distribution

$$\mathbb{P}(f(r) = v) = \frac{d(v)}{2|E(G)|}.$$

Next, for each child x of the root, we let $f(x)$ be a uniformly random neighbour of $f(r)$, where all of these choices are made independently of one another. Generally speaking, for $x \in V(T) \setminus \{r\}$, we assume that $f(p(x))$ has been chosen already and let $f(x)$ be a uniformly random neighbour of $f(p(x))$ chosen independently of all previous choices conditional on the choice of $f(p(x))$. The fact that f is a homomorphism is clear from construction.

Now, let us show that f has the desired properties. The first property clearly holds for $x = r$ by construction. For $x \in V(T) \setminus \{r\}$ and $v \in V(G)$, we have

$$\mathbb{P}(f(x) = v) = \sum_{u \in N(v)} \mathbb{P}(f(x) = v \mid f(p(x)) = u)\mathbb{P}(f(p(x)) = u)$$

which, by induction on the distance from x to y and construction of f, is equal to

$$\sum_{u \in N(v)} \frac{1}{d(u)} \cdot \frac{d(u)}{2|E(G)|} = \frac{d(v)}{2|E(G)|}.$$

Finally, let $xy \in E(T)$ and $uv \in E(G)$. If x is the parent of y, then $\mathbb{P}(f(y) = v \mid f(x) = u) = \frac{1}{d(u)}$ simply by construction. So, we assume that y is the parent of x. We have

$$\mathbb{P}(f(y) = v \mid f(x) = u) = \frac{\mathbb{P}(f(y) = v \text{ and } f(x) = u)}{\mathbb{P}(f(x) = u)} = \frac{\mathbb{P}(f(x) = u \mid f(y) = v)\mathbb{P}(f(y) = v)}{\mathbb{P}(f(x) = u)}$$

$$= \frac{(1/d(v)) \cdot (d(v)/2|E(G)|)}{(d(u)/2|E(G)|)} = \frac{1}{d(u)}.$$

This completes the proof. □

Finally, we prove the theorem.

Proof of Theorem 6.44 ([170, 250]). Let $n = |V(G)|$. If T has only one vertex (and therefore no edges), then the result is trivial.[3] So, assume that T has at least two vertices. Let r be an arbitrary root vertex of T and define "parents" and "children" as in the proof of Proposition 6.46.

Let T' be a graph obtained from T by adding $|V(T)| - 2$ isolated vertices. Note that

$$\hom(T', G) = n^{|V(T)|-2} \hom(T, G). \tag{6.47}$$

Consider a random homomorphism f from T' to G where each of the vertices of T are mapped according to the distribution described in Proposition 6.46 and each isolated vertex is mapped to any given vertex v of G with probability $\frac{d(v)}{2|E(G)|}$ independently of all other vertices. Then, by Lemma 6.30 (a), we have

$$\log(\hom(T', G)) \geq H(f) = H(f(x) : x \in V(T')).$$

Label the vertices of T' by $v_1, \ldots, v_{2|V(T)|-2}$ in such a way that the first $|V(T)| - 1$ vertices are the isolated vertices and the root r of T, and if $v_i \in V(T) \setminus \{r\}$, then $p(v_i)$ comes before v_i in the list. Then

$$H(f) = H(f(v_1), \ldots, f(v_{2|V(T)|-2})) = \sum_{i=1}^{2|V(T)|-2} H(f(v_i) \mid f(v_1), \ldots, f(v_{i-1})).$$

Since the image of each vertex is independent of the image of all vertices that came before it, except for its parent (if it has one), this can be rewritten as

$$\sum_{i=1}^{|V(T)|-1} H(f(v_i)) + \sum_{i=|V(T)|}^{2|V(T)|-2} H(f(v_i) \mid f(p(v_i))).$$

[3]In the case that $|E(G)| = 0$, we view 0^0 as being 1.

Let $V(K_2) = \{1, 2\}$ and let g be a uniformly random homomorphism from K_2 to G. By Claim 6.45 and Proposition 6.46, we have that $f(v_i)$ has the same distribution as $g(1)$. Also, for any $i \geq |V(T)|$, the distribution of $f(v_i)$ given $f(p(v_i))$ is also identical to the distribution of $g(2)$ given $g(1)$. If two random variables have the same distribution, then they have the same entropy. Thus, this expression can be rewritten as

$$(|V(T)| - 1)H(g(1)) + (|V(T)| - 1)H(g(2) \mid g(1)) = |E(T)| \cdot H(g(1), g(2))$$
$$= |E(T)| \cdot H(g) = |E(T)| \cdot \log(\hom(K_2, G)).$$

Therefore,

$$\hom(T', G) \geq \hom(K_2, G)^{|E(G)|}.$$

Now, putting everything together, we get

$$t(T, G) = \frac{\hom(T, G)}{n^{|V(T)|}} = \frac{\hom(T', G)}{n^{2|V(T)|-2}} \geq \frac{\hom(K_2, G)^{|E(T)|}}{n^{2|E(T)|}} = t(K_2, G)^{|E(T)|}$$

and so we are done. □

Note that Theorem 6.44 is best possible (asymptotically) by letting G be a large random graph. In fact, any d-regular graph would also work; see Exercise 6.19.

6.6 Exercises

6.1 Suppose that G is a graph on n vertices such that every 6 vertices of G contains an independent set of size 3. Prove that

$$\alpha(G) \geq (1 + o(1))\sqrt{\frac{n \ln(n)}{2}}.$$

Hint: It is easier than it looks, if you use Theorem 6.3 in the right way. Think about the triangles.

6.2 Prove that there exists a constant $c > 0$ such that every K_4-free graph G on n vertices satisfies

$$\alpha(G) \geq (c + o(1)) \cdot (n \ln(n))^{1/3}.$$

6.3 Prove that the chromatic number of a triangle-free graph on n vertices is $O(\sqrt{n})$.

6.4 For $1 \leq d \leq n - 1$ such that nd is even, let G be a d-regular graph with n vertices.

(a) Prove that every independent set in G has cardinality at most $n/2$. Describe a large family of tight examples (i.e. for infinitely different values of d and n).

(b) Show that if G has an independent set of cardinality at least $n/2 - t$, then there is a set E' of at most td edges such that $G \setminus E'$ is bipartite.

(c) Show that the number of edges of G contained in any set of at least $n/2 + t$ vertices is at least td.

6.5 Let $d \geq 1$, let n be divisible by $2d$ and let G be a disjoint union of $n/(2d)$ copies of $K_{d,d}$.

(a) How many proper 2-colourings does $K_{d,d}$ have? How many proper 2-colourings does G have?

(b) How many proper 3-colourings does $K_{d,d}$ have? (Careful!) How many proper 3-colourings does G have?

6.6 Let G be a d-regular bipartite graph with bipartition (A, B) where $|A| = |B| = n$.

(a) For $0 \leq k \leq n$, prove that the number of independent sets S of G such that $|S \cap A| = k$ is at least

$$\binom{n}{k} 2^{n-kd}.$$

(b) Suppose now that $n = 2^{d-1}$. Let k be fixed. Argue that, for large d,

$$\binom{2^{d-1}}{k} 2^{2^{d-1}-kd} = (1 + o(1)) \frac{\left(\frac{1}{2}\right)^k}{k!} 2^{2^{d-1}}.$$

(c) Prove that the number of independent sets in the d-dimensional hypercube Q_d is at least

$$(1 + o(1)) 2\sqrt{e} 2^{2^{d-1}}$$

for large d.

(d) Using similar ideas, prove that the number of proper 3-colourings of Q_d is at least

$$(1 + o(1)) 6e 2^{2^{d-1}}.$$

6.7 (a) Prove that the number of triangle-free graphs with vertex set $[n]$ is at least

$$2^{\lfloor n^2/4 \rfloor}.$$

(b) Consider the following lemma of Balogh, Morris and Samotij:

Lemma: For every $\varepsilon > 0$ there exists $n_0(\varepsilon)$ such that if $n \geq n_0(\varepsilon)$, then there exists a collection $\mathcal{G}$ of graphs with vertex set $[n]$ such that

- $|\mathcal{G}| \leq 2^{\varepsilon n^2}$,
- every graph in $\mathcal{G}$ has at most $(1/4 + \varepsilon)n^2$ edges and
- for every triangle-free graph G with vertex set $[n]$, there exists $G' \in \mathcal{G}$ such that $G \subseteq G'$.

Use this lemma to prove that the number of triangle-free graphs on n vertices is at most

$$2^{n^2/4 + o(n^2)}.$$

6.8 Let $q \geq 2$, let G be a d-regular graph on n vertices and let m be a positive integer. For each independent set S of G, let F_S and C_S be defined as in the proof of Theorem 6.9 with respect to G and m. Also, let g be the mapping as in the proof of Theorem 6.9.

(a) Show that, if f is a proper q-colouring of G, then

$$\bigcup_{c=1}^{q} \left(F_{f^{-1}(c)} \cup C_{f^{-1}(c)}\right) = V(G).$$

(b) Say that the *fingerprint* of a proper q-colouring f of G is the vector of sets $(F_1, \dots, F_q)$ such that F_c is the fingerprint of $F_{f^{-1}(c)}$ for all $1 \leq c \leq q$. Given a vector $(F_1, \dots, F_q)$ of subsets of $V(G)$ of cardinality at most n/m and a vertex $v \in V(G)$, let $\eta_{F_1,\dots,F_q}(v)$ be the number of c such that $1 \leq c \leq q$ and $v \in g(F_c)$.

Prove that the number of proper q-colourings with fingerprint $(F_1, \dots, F_q)$ is at most

$$\prod_{v \in V(G)} \eta_{F_1,\dots,F_q}(v).$$

(c) Use the AM-GM Inequality (look it up if you don't know it) to show that the number of proper q-colourings with fingerprint $(F_1, \dots, F_q)$ is at most

$$\left(\sum_{v \in V(G)} \frac{\eta_{F_1,\dots,F_q}(v)}{n}\right)^n$$

and then argue that this quantity is equal to

$$\left(\sum_{c=1}^{q} \frac{|C_c|}{n}\right)^n.$$

(d) Using (6.12), (6.13) and the previous parts of the exercise, show that the number of proper q-colourings of G is at most

$$\left[(n/m+1)^q (em)^{nq/m}\right] q^n \left(\frac{1}{2} + \frac{m}{2d}\right)^n.$$

(e) Complete the proof of Theorem 6.17.

Hint 1: Let $\varepsilon > 0$ and choose m large enough with respect to q and ε so that the inequality $(em)^q < \left(1 + \frac{2\varepsilon}{3}\right)^m$ holds.

Hint 2: Now, choose d large enough with respect to m so that $m/2d$ is less than $\varepsilon/3$.

6.9 The *matching polynomial* of a graph G is defined to be

$$M_G(\lambda) = \sum_{M \in \mathcal{M}(G)} \lambda^{|M|}$$

where $\mathcal{M}(G)$ is the set of all matchings of G.

(a) For $0 \leq k \leq d$, what is the coefficient of λ^k in $M_{K_{d,d}}(\lambda)$?

(b) Let X be a random matching of G chosen according to the distribution

$$\mathbb{P}(X = M) = \frac{\lambda^{|M|}}{M_G(\lambda)}.$$

Prove that the expected cardinality of X is

$$\sum_{e \in E(G)} \mathbb{P}(e \in X) = \lambda \cdot \frac{\mathrm{d}}{\mathrm{d}\lambda} \ln(M_G(\lambda)).$$

6.10 Let X be a random variable with range $\mathbb{N}$ such that, for all $n \geq 1$,

$$\mathbb{P}(X = n) = \frac{1}{2^n}.$$

Determine the entropy of X.

6.11 Let $k_1, \ldots, k_n$ be integers such that

$$\sum_{i=1}^{n} \left(\frac{1}{2}\right)^{k_i} = 1.$$

For $1 \leq i \leq n$, let $s_i \in \{0,1\}^{k_i}$. Suppose that X is a random variable such that $\mathbb{P}(X = i) = (1/2)^{k_i}$. That is, s_i is a binary string of length k_i. Prove that the entropy of X is equal to the expected value of the length of s_X.

6.12 Prove property (c) of Lemma 6.30.

Note that you can either do this from first principles, or you can apply the result of Example 6.28 and property (b) of Lemma 6.30.

6.13 Prove property (d) of Lemma 6.30.

Note that the case when X and Y are independent was already covered in Example 6.27. What remains to be shown is that $H(X \mid Y) < H(X)$ whenever X and Y are not independent.

6.14 The *binary entropy function* $h : [0,1] \to \mathbb{R}$ is defined by

$$h(p) := \begin{cases} -p\log_2(p) - (1-p)\log_2(1-p) & \text{if } p \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{F} \subseteq 2^{[n]}$ be a set system. For each $i \in [n]$, let

$$p_i = \frac{|\{A \in \mathcal{F} : i \in A\}|}{|\mathcal{F}|}.$$

Prove that

$$|\mathcal{F}| \leq 2^{\sum_{i=1}^{n} h(p_i)}.$$

6.15 A family $\mathcal{F} \subseteq 2^{[n]}$ is *union-closed* if, whenever $A, B \in \mathcal{F}$, we also have $A \cup B \in \mathcal{F}$. Our goal is to use a lemma about entropy (that we will not prove) to show that, for any union-closed family $\mathcal{F} \neq \{\emptyset\}$, there exists $i \in [n]$ which is contained in a large number of sets of $\mathcal{F}$.

(a) Let $h(p)$ be the binary entropy function defined as in Exercise 6.14. Let $p \in (0,1)$ and suppose that A is a random subset of $[n]$ obtained by including each $i \in [n]$ in A with probability p independently of one another. Determine $H(A)$ in terms of $h(p)$ and n.

(b) Suppose that A and B are two sets chosen independently of one another with the distribution described in part (a). Determine $H(A \cap B)$ and $H(A \cup B)$.

(c) Consider the following facts about the binary entropy function:

Fact 1: $h(p)$ is increasing for $0 \leq p \leq 1/2$ and decreasing for $1/2 \leq p \leq 1$.

Fact 2: $h(p) = h(q)$ if and only if $p = q$ or $p = 1 - q$.

Using these two facts and the previous parts of the question, show that, if $p \in (0,1)$ and A and B are two sets chosen independently of one another with the distribution described in part (a), then

$$H(A \cup B) \begin{cases} > H(A) & \text{if } p < \frac{3-\sqrt{5}}{2} \\ = H(A) & \text{if } p = \frac{3-\sqrt{5}}{2}, \\ < H(A) & \text{if } p > \frac{3-\sqrt{5}}{2}. \end{cases}$$

(d) Consider the following lemma of [229]:

Lemma: If $p < \frac{3-\sqrt{5}}{2}$ and A and B are independently chosen according to some distribution on $2^{[n]}$ such that $\mathbb{P}(A = \emptyset) \neq 1$ and $\mathbb{P}(i \in A) \leq p$ for all $i \in [n]$, then

$$H(A \cup B) > H(A).$$

Explain how the previous part of the question shows that this lemma is tight.

(e) Use the lemma in the previous part of the question to prove that every union-closed family $\mathcal{F} \neq \{\emptyset\}$ there exists $i \in [n]$ such that i is contained in at least $\left(\frac{3-\sqrt{5}}{2}\right)|\mathcal{F}|$ elements of $\mathcal{F}$.

Hint: Lemma 6.30 (a) is useful.

6.16 Our aim in this exercise is to derive Theorem 6.43 from Corollary 6.42. Let G be a graph and let A be its adjacency matrix.

(a) Let $\mathcal{H}$ be the collection of all subgraphs of H of G such that $V(H) = V(G)$ and H is a disjoint union of cycles and edges such that every vertex has degree at least one. For each $H \in \mathcal{H}$, let $s(H)$ be the number of components of H with more than two vertices. Show that

$$\text{per}(A) = \sum_{H \in \mathcal{H}} 2^{s(H)}.$$

(b) Show that

$$\text{pm}(G)^2 = \sum_{\substack{H \in \mathcal{H} \\ H \text{ is bipartite}}} 2^{s(H)}.$$

(c) Prove Theorem 6.43 using Corollary 6.42.

6.17 A *Hamiltonian cycle* in a graph G is a cycle in G which contains every vertex of G. Prove that, if G is a bipartite graph with bipartition (A, B), then the number of Hamiltonian cycles in G is at most

$$\prod_{v \in A} (d(v)!)^{2/d(v)}.$$

6.18 Prove that the number of cycles in any graph G is at most

$$\prod_{v \in V(G)} (d(v)+1)!^{1/(d(v)+1)}.$$

Hint: Consider the permanent of the matrix $A + I$ where A is the adjacency matrix of G and I is the identity.

6.19 A graph is *d-regular* if all of its vertices have degree equal to d. Compute $t(T, G)$ for any tree T on k vertices and any d-regular graph G on n vertices. Rewrite this as $t(K_2, G)$ raised to some power.

6.20 Given a graph H, the 2-*blowup* of $H(2)$ of H is the graph with vertex set $V(H) \times \{1, 2\}$ where, for $u, v \in V(H)$ and $i, j \in \{1, 2\}$, the vertices (u, i) and (v, j) are adjacent if and only if $uv \in E(H)$. Using a similar argument to that in the proof of Theorem 6.44, prove that, if T is a tree, then

$$t(T(2), G) \geq t(K_2, G)^{|E(T(2))|}$$

for every graph G.

Hint: Use Theorems 4.8 and 6.44.

6.21 Suppose that H and F are graphs such that

$$t(H, G)^{|E(F)|} \geq t(F, G)^{|E(H)|} \tag{6.48}$$

for every graph G. Prove that

$$\frac{|E(H)|}{|V(H)|} \geq \frac{|E(F)|}{|V(F)|}.$$

Hint: Apply (6.48) for an appropriately chosen graph G.

6.22 Let F and H be graphs. Suppose that there exists $\alpha > 0$ such that the following holds for every graph G:

$$t(H, G) \geq \frac{t(F, G)^{\alpha}}{1867}.$$

Prove that, in fact, the following stronger inequality must hold for every graph G:

$$t(H, G) \geq t(F, G)^{\alpha}.$$

Hint: Suppose there is a graph G such that $t(H, G) < t(F, G)^{\alpha}$. Use the result of Exercise 4.10.

6.23 Our goal is to give a proof of the following theorem of Conlon, Fox and Sudakov [51]:

Theorem: If H is a bipartite graph with bipartition (A, B) and there is a vertex $x \in A$ such that $N(x) = B$, then, for every graph G,

$$t(H, G) \geq t(K_2, G)^{|E(H)|}.$$

Let $f_x : \{x\} \cup B \to V(G)$ be a random function sampled as follows:

- $f_x(x)$ is a random vertex of G chosen with probability proportional to its degree, as in the proof of Proposition 6.46.
- for each $u \in B$, $f_x(u)$ is a uniformly random neighbour of $f_x(x)$.

Given $z \in A \setminus \{x\}$, let $f_z : \{z\} \cup N(z) \to V(G)$ be a random function chosen in the following way. First, sample $f_z : \{z\} \cup B \to V(G)$ according to the same distribution as f_x, but with z in the place of x, independently of f_x. While there exist $y \in N(z)$ such that $f_z(y) \neq f_x(y)$, resample f_z in the same way. Once $f_z(y) = f_x(y)$ for all $y \in N(z)$, stop and restrict the domain of the f_z to $\{z\} \cup N(z)$.

(a) For and $z \in A \setminus \{x\}$, vertex $v \in V(G)$ and tuple $(u_y : y \in N(v)) \in V(G)^{N(v)}$, prove that

$$\mathbb{P}(f_z(z) = v \mid f_z(y) = u_y \text{ for all } y \in N(z)) = \mathbb{P}(f_x(x) = v \mid f_x(y) = u_y \text{ for all } y \in N(z)).$$

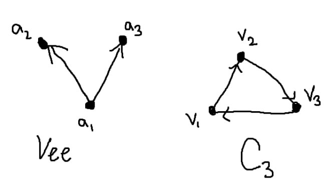
(b) Let H' be a graph obtained from H by adding $|E(H)| - |V(H)|$ isolated vertices. Find a distribution on the set of homomorphisms from H' to G with entropy at least $|E(H)| \cdot \log(\hom(K_2, G))$.

(c) Prove the theorem.

6.24 A *digraph* is a pair (V, A) where V is a set of vertices and A is a set of ordered pairs of vertices called *arcs*. That is, $A \subseteq V^2$. The way that we depict a digraph is by drawing its vertices as points and drawing an arrow from vertex u to vertex v if $(u, v) \in A$.

A *homomorphism* from a digraph H to a digraph D is a mapping $f : V(H) \to V(D)$ such that $(f(u), f(v)) \in A(D)$ whenever $(u, v) \in A(H)$. We define $\mathrm{Hom}(H, D)$, $\hom(H, D)$ and $t(H, D)$ analogously to the case of graphs.

Let Vee and C_3 be the digraphs depicted below.



The goal of this problem is to show that every digraph D satisfies $\hom(Vee, D) \geq \hom(C_3, D)$ and that this inequality is tight.

(a) For each $n \geq 1$, let D_n be the digraph with vertex set $V_1 \cup V_2 \cup V_3$ such that $|V_i| = n$ for each $1 \leq i \leq 3$ and $(u,v) \in A(D_n)$ if and only if $(u,v) \in V_1 \times V_2$, $(u,v) \in V_2 \times V_3$ or $(u,v) \in V_3 \times V_1$. Prove that

$$\hom(Vee, D_n) = \hom(C_3, D_n) = 3n^3.$$

(b) Let D be any digraph such that $\hom(C_3, D) \geq 1$. Let F be a uniformly random element of $\mathrm{Hom}(C_3, D)$. For $i \in \{1,2,3\}$, let X_i be the random variable $F(v_i)$. Using properties (b) and (d) of Lemma 6.30, prove that

$$H(F) = H(X_1, X_2, X_3) \leq H(X_1) + 2H(X_2 \mid X_1).$$

(c) Let Y_1, Y_2, Y_3 be random variables defined by setting

$$\mathbb{P}(Y_1 = v) = \mathbb{P}(X_1 = v)$$

for each $v \in V(D)$ and then, for each $v, w \in V(D)^2$, and $i \in \{2,3\}$, setting

$$\mathbb{P}(Y_i = w \mid Y_1 = v) = \mathbb{P}(X_2 = w \mid X_1 = v).$$

Prove that

$$H(Y_1, Y_2, Y_3) = H(X_1) + 2H(X_2 \mid X_1).$$

(d) Using the previous two parts of the question, along with the right properties of entropy, prove that

$$\hom(Vee, D) \geq \hom(C_3, D)$$

for every digraph D. Conclude that $t(Vee, D) \geq t(C_3, D)$ holds for every digraph as well.

6.7 Challenge Problems

6.1* A *Latin square* of order n is an $n \times n$ matrix in which each of the symbols in $\{1, \ldots, n\}$ appears exactly once in every row and column (sort of like Sudoku, but with one fewer restriction). Let $L(n)$ be the number of Latin squares of order n. Prove that

$$L(n)^{1/n^2} = (1 + o(1))\frac{n}{e^2}.$$

Hint: Look up van der Waerden's Conjecture (which was proved in [71, 93]) for lower bounding the permanent of a "doubly stochastic matrix" and apply it.

6.2* Recall that P_k is the path with k vertices. Prove that, for every graph G,

$$t(P_6, G)^3 \geq t(P_4, G)^5.$$

Generalize this to prove that, for all $k \geq 2$,

$$t(P_{k+2}, G)^{|E(P_k)|} \geq t(P_k, G)^{|E(P_{k+2})|}.$$

Hint: Construct a distribution on homomorphisms from P_{k+2} to G and play around with its entropy.

Chapter 7

Graph Limits

The purpose of this chapter is to provide a very surface-level introduction to the theory of graph limits. This chapter has been heavily influenced by the excellent monograph of Lovász [184]; we refer the reader to [184] for a more in-depth treatment of the subject.

7.1 Optimal Constructions for Extremal Graph Theory Problems

The focus in this chapter is on a particular notion of "convergence" for sequences of finite graphs and the "completion" of the set of finite graphs under this notion of convergence. While this may sound unusual at first, it is fairly similar to the construction of the real numbers from the rationals.[1] Rational numbers, like graphs, are fundamentally discrete objects. Consider the following optimization problem:

$$\begin{aligned} &\text{Minimize } x^3 - 6x \\ &\text{subject to } 0 \leq x \leq 2. \end{aligned} \tag{7.1}$$

Any first-year Calculus student will tell you that the minimum occurs at either $x = 0$, $x = 2$ or a critical point of $x^3 - 6x$. The only critical point in $[0, 2]$ is $x = \sqrt{2}$, which turns out to be optimal. Of course, the number $\sqrt{2}$ is famously irrational. Thus, while rational numbers are wonderful discrete objects which are easy to conceptualize, they are insufficient for representing the solutions to natural optimization problems expressed in terms of rational numbers themselves. However, while the optimal choice of x is not rational, there does exist a sequence $x_1, x_2, \dots$ of rational numbers in $[0, 2]$ such that $x_n^3 - 6x_n$ converges to the optimum value as n tends to infinity. Thus, the optimum cannot be achieved by a rational number but it can be approximated.

Now, let us observe a similar issue that crops up in extremal graph theory. Consider the following optimization problem for a graph G:

$$\begin{aligned} &\text{Minimize } t(C_4, G) \\ &\text{subject to } 1/2 \leq t(K_2, G) \leq 1. \end{aligned} \tag{7.2}$$

In Theorem 4.8, we showed that $t(C_4, G) \geq t(K_2, G)^4$ for every graph G. Therefore, the solution to (7.2) is at least $1/16$. Moreover, as we discussed at the end of Section 4.2, there exists a sequence $G_1, G_2, \dots$ of graphs such that $t(K_2, G_n) \to 1/2$ and $t(C_4, G_n) \to 1/16$ as $n \to \infty$. Given this,

[1] This lovely analogy is borrowed from [270, Chapter 4].

it is natural to wonder: does there exist a single finite graph G such that $t(K_2, G) = 1/2$ and $t(C_4, G) = 1/16$? By inspecting the proof of Theorem 4.8 carefully, one will find that, in fact, no such graph exists (see Exercise 7.1). Thus, (7.2) has no solution in the set of finite graphs, just as (7.1) has no solution in the set of rational numbers. This motivates us to seek an extension of the set of finite graphs to a richer continuous space.

Before providing the full "completion" of the set of finite graphs, let us take an intermediate step to finite matrices. Given a graph G with vertex set $[n]$, let A_G be its adjacency matrix. Given a matrix A, let $A(i,j)$ be the entry in the ith row and jth column of A. Then, for any graph H, $\hom(H, G)$ can be conveniently rewritten as follows:

$$\hom(H, G) = \sum_{f:V(H)\to[n]} \left(\prod_{uv\in E(H)} A_G(f(u), f(v)) \right).$$

The above equation holds due to the fact that the product within the sum evaluates to one if f is a homomorphism and zero otherwise. The following is a natural extension to general matrices.

Definition 7.3. Given a graph H and an $n \times n$ matrix A over $\mathbb{R}$, define

$$\hom(H, A) := \sum_{f:V(H)\to[n]} \left(\prod_{uv\in E(H)} A(f(u), f(v)) \right)$$

and

$$t(H, A) := \frac{\hom(H, A)}{n^{|V(H)|}}.$$

By equating a graph with its adjacency matrix, we see that the set of $[0,1]$-valued squared matrices A which are *symmetric*, in the sense that $A = A^T$, can be seen as a reasonable extension of the set of finite graphs. Suppose that we extend (7.2) to the following optimization problem for such matrices:

$$\begin{aligned} &\text{Minimize } t(C_4, A) \\ &\text{subject to } 1/2 \le t(K_2, A) \le 1. \end{aligned} \tag{7.4}$$

By generalizing the proof of Theorem 4.8, one can also obtain a lower bound of $1/16$ for this problem. However, in contrast to (7.2), the minimum for this problem can be achieved. Indeed, suppose that we let A be the $n \times n$ matrix in which every entry is equal to $1/2$. Then, by Definition 7.3, we have $t(K_2, A) = 1/2$ and $t(C_4, A) = 1/16$. Thus, the set of $[0,1]$-valued symmetric matrices is sufficiently rich for the optimum of (7.4) to be achieved but it is not so large that the optimal value changes.

7.2 Convergence in Homomorphism Density

Unfortunately, $n\times n$ matrices are still too "discrete" to be able to capture the optimal constructions for all of the types of extremal problems that we are interested in. Intuitively speaking, what is needed is to let n "go to infinity." Consider the following few definitions.

Definition 7.5. A function U with domain $[0,1]^2$ is said to be *symmetric* if $U(x,y) = U(y,x)$ for all $x, y \in [0,1]$.

Definition 7.6. A *kernel* is a bounded symmetric measurable function $U : [0,1]^2 \to \mathbb{R}$.

The way that one should think of a kernel is as a real symmetric matrix with uncountably many rows and columns. The condition that U is measurable is only required so that various integrals involving U are well-defined (see, e.g., Definition 7.9 below).

Definition 7.7. A *graphon* is a kernel W such that $0 \leq W(x, y) \leq 1$ for all $x, y \in [0, 1]$.

Example 7.8. Let A be an $n \times n$ symmetric matrix and let $I_1, \ldots, I_n$ be a partition of $[0, 1]$ into intervals of length $1/n$. Define $U_A : [0, 1]^2 \to \mathbb{R}$ to be the function defined by $U_A(x, y) = A(i, j)$ whenever $x \in I_i$ and $y \in I_j$. Then U_A is a *kernel representation* of A. Given a graph G, let $W_G = U_{A_G}$. Then W_G is a *graphon representation* of G.

Next, we extend the notion of homomorphism density to kernels. Since kernels are analytic objects, the sum in Definition 7.3 is replaced by an integral.

Definition 7.9. Given a graph H and a kernel U, the *homomorphism density* of H in U is defined to be

$$t(H, U) := \int_{[0,1]^{V(H)}} \prod_{uv \in E(H)} U(x_u, x_v) dx_{V(H)}$$

where $x_{V(H)} = (x_v : v \in V(H))$ is a vector indexed by the vertices of H.

The following simple observation tells us that the homomorphism density for kernels extends that of matrices which, itself, is a generalization of the notion for graphs.

Observation 7.10. *For any symmetric square matrix A and graph H,*

$$t(H, A) = t(H, U_A).$$

Finally, we introduce a fundamental notion of convergence for sequences of graphs.

Definition 7.11. A sequence $(U_n)_{n=1}^{\infty}$ of kernels is said to be *left-convergent* if $(t(H, U_n))_{n=1}^{\infty}$ converges for every graph H. Moreover, given a kernel U, we say that the sequence *converges to U* if

$$\lim_{n \to \infty} t(H, U_n) = t(H, U)$$

for every graph H.

Definition 7.12. Given a sequence $(G_n)_{n=1}^{\infty}$ of graphs and a graphon W, we say that $(G_n)_{n=1}^{\infty}$ *left-converges (to W)* if the sequence $(W_{G_n})_{n=1}^{\infty}$ left-converges (to W).

Let us consider a few examples to illustrate this notion of convergence.

Example 7.13. Let $(G_n)_{n=1}^{\infty}$ be the sequence defined by $G_n = K_{n,n}$ for all $n \geq 1$. We show that the sequence $(G_n)_{n=1}^{\infty}$ is left-convergent. If H is non-bipartite, then $\hom(H, G_n) = 0$ and so $t(H, G_n) = 0$ for all $n \geq 1$. Next, suppose that H is bipartite. First, suppose that H is connected. If the bipartition of H is (A, B), then the number of homomorphisms from H to G_n is equal to the number of functions from $V(H)$ to $V(G_n)$ such that all vertices of A map to one side of the bipartition and all vertices of B map to the other. Therefore,

$$\hom(H, G_n) = 2 \left(\frac{n}{2}\right)^{|A|} \left(\frac{n}{2}\right)^{|B|},$$

So,

$$t(H, G_n) = \left(\frac{1}{2}\right)^{|A|+|B|-1} = \left(\frac{1}{2}\right)^{|V(H)|-1}.$$

Next, if H is disconnected, then $H = H_1 \sqcup H_2 \sqcup \cdots \sqcup H_k$ for some connected graphs $H_1, \ldots, H_k$. Therefore, by Exercise 4.9 and the fact that $H_1, \ldots, H_k$ are connected,

$$t(H, G_n) = \prod_{i=1}^{k} t(H_i, G_n) = \prod_{i=1}^{k} \left(\frac{1}{2}\right)^{|V(H_i)|-1} = \left(\frac{1}{2}\right)^{|V(H)|-k}.$$

Thus, in this example, we that the sequence $(t(H, G_n))_{n=1}^{\infty}$ is constant for every graph H. Therefore, $G_1, G_2, \ldots$ is left-convergent. It left-converges to W_{K_2}.

Example 7.14. Consider the random graph $G(n, 1/2)$. For any graph H, we have

$$\mathbb{E}(t(H, G(n, 1/2))) = (1 + o(1)) \left(\frac{1}{2}\right)^{|E(H)|}$$

as $n \to \infty$. Moreover, the value of $t(H, G(n, 1/2))$ is "concentrated" around its expected value. Therefore, for each graph H, the sequence $(t(H, G(n, 1/2)))_{n=1}^{\infty}$ converges with probability one. Therefore, the sequence $(G(n, 1/2))_{n=1}^{\infty}$ is left-convergent with probability 1.

The main goal in this chapter is to prove the following magnificent result of Lovász and Szegedy [187], which we refer to as the "Lovász–Szegedy Compactness Theorem."

Theorem 7.15 (Lovász and Szegedy [187])**.** *For any sequence $(W_n)_{n=1}^{\infty}$ of graphons, there exists a W and a subsequence $(W_{n_k})_{k=1}^{\infty}$ such that the subsequence $(W_{n_k})_{k=1}^{\infty}$ is left-convergent to W. Moreover, for any graphon W there exists a sequence $(G_n)_{n=1}^{\infty}$ of finite graphs with $|V(G_n)| \to \infty$ which left-converges to W.*

Let us now connect this theorem to the types of optimization problems discussed in Section 7.1. Let $k, \ell \geq 1$, let $H_1, \ldots, H_k$ be graphs and let $c_i, b_{i,j}, d_j \in \mathbb{R}$ for all $i \in [k]$ and $j \in [\ell]$. Consider the optimization problem for a graphon W:

$$\begin{aligned} \text{Minimize } & \sum_{i=1}^{k} c_i \cdot t(H_i, W) \\ \text{subject to } & \sum_{i=1}^{k} b_{i,j} \cdot t(H_i, W) \geq d_j \text{ for all } j \in [\ell]. \end{aligned} \tag{7.16}$$

This generic formulation captures a wide range of extremal problems. Now, assuming that the constraints in (7.16) are feasible, Theorem 7.15 immediately tells us that the minimum is actually achieved. That is, if we let $W_1, W_2, \ldots$ be a sequence of graphons satisfying the constraints in (7.16) such that $\sum_{i=1}^{k} c_i \cdot t(H_i, W_n)$ converges to the optimal value as $n \to \infty$, then we can take W to be the limit of a subsequence of $W_1, W_2, \ldots$. By definition of left-convergence, we have that W satisfies all constraints in (7.16) and, subject to this, minimizes the objective function. Using the "moreover" part of Theorem 7.15, we get that there is a sequence $G_1, G_2, \ldots$ of graphs such that

$$\sum_{i=1}^{k} b_{i,j} \cdot t(H_i, G_n) \geq d_j - o(1) \text{ for all } j \in [\ell]$$

and

$$\sum_{i=1}^{k} c_i \cdot t(H_i, G_n) \to \sum_{i=1}^{k} c_i \cdot t(H_i, W).$$

Therefore, the optimum is asymptotically attained by a sequence of finite graphs. In this sense, graphons and graphs are truly analogous to real and rational numbers; real numbers complete the rational numbers, and the rationals are dense in the reals. Our focus in the rest of this chapter is on proving Theorem 7.15. To do so, we will need to build up a few more ideas.

7.3 Norms and Metrics for Graphons

In analysis, many of the common notions of convergence of functions are described in terms of norms or metrics. As it turns out, left-convergence of graphs can also be described in terms of a norm; however, the important norm for graph limits (and therefore, for extremal graph theory) is *not* one of the typical norms studied in an analysis class.

Definition 7.17 (The Cut Norm for Kernels)**.** The *cut norm* of a kernel U is defined to be

$$\|U\|_\square := \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} U(x,y)dxdy \right|.$$

The fact that the cut norm is, indeed, a norm is not hard to show; see Exercise 7.2. In order to gain some intuition about the cut norm, let us spend some time to compare it to some of the norms that may be more familiar to you.

Definition 7.18. Let U be a kernel. For $p \in [1, \infty)$, define

$$\|U\|_p := \left(\int_{[0,1]^2} |U(x,y)|^p dxdy \right)^{1/p}.$$

For $p \in [1, \infty)$, we refer to $\|U\|_p$ as the ℓ_p*-norm* of U.

Definition 7.19. Given a kernel U, let

$$\|U\|_\infty := \sup_{x,y\in[0,1]} |U(x,y)|.$$

We refer to $\|U\|_\infty$ as the ℓ_∞*-norm* of U. Note that this supremum exists as kernels are bounded.

Let us recall the following standard fact about the ℓ_p-norms (which follows from Hölder's Inequality).

Proposition 7.20. *If* $1 \le p \le q \le \infty$*, then* $\|U\|_p \le \|U\|_q$ *for any kernel* U*.*

Proposition 7.20 will be derived from the following lemma regarding finite matrices.

Lemma 7.21. *Let* A *be an* $n \times n$ *real matrix and let* $1 \le p < q$ *be real numbers. Then*

$$\left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |A(i,j)|^p \right)^{1/p} \le \left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} |A(i,j)|^q \right)^{1/q}.$$

Proof. Let $s = \frac{q}{p}$ and note that $s > 1$. By Corollary C.6,

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|A(i,j)|^p = \sum_{i=1}^{n}\sum_{j=1}^{n}\left|\frac{|A(i,j)|^p}{n^2}\right| \leq n^{\frac{2(s-1)}{s}}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\left|\frac{|A(i,j)|^p}{n^2}\right|^s\right)^{1/s}$$

$$= \left(n^{2(s-1)}n^{-2s}\sum_{i=1}^{n}\sum_{j=1}^{n}|A(i,j)|^q\right)^{p/q} = \left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|A(i,j)|^q\right)^{p/q}.$$

The result follows by taking the pth root of both sides. □

Proof of Proposition 7.20. If $p = q$, then there is nothing to prove; so, we assume that $q > p$. Now, if $q = \infty$, then

$$\|U\|_p = \left(\int_{[0,1]^2}|U(x,y)|^p dxdy\right)^{1/p} \leq \left(\int_{[0,1]^2}\|U\|_\infty^p\right)^{1/p} = \|U\|_\infty.$$

So, we can assume that $1 \leq p < q < \infty$. For each $n \geq 1$, we let A_n be a $2^n \times 2^n$ matrix in which, for $1 \leq i, j \leq 2^n$, we have

$$A_n(i,j) := 2^{2n}\int_{\left[\frac{i-1}{2^n},\frac{i}{2^n}\right)}\int_{\left[\frac{j-1}{2^n},\frac{j}{2^n}\right)} U(x,y)dxdy.$$

In other words, we divide $[0,1]^2$ into a $2^n \times 2^n$ grid and, for $1 \leq i, j \leq 2^n$, we set $A_n(i,j)$ to be the average of $U(x,y)$ over all (x,y) in the (i,j) cell of the grid. Standard facts from analysis (e.g. [184, Proposition 9.8]) tell us that

$$\|U\|_p = \lim_{n\to\infty}\left(\frac{1}{2^{2n}}\sum_{i=1}^{2^n}\sum_{j=1}^{2^n}|A_n(i,j)|^p\right)^{1/p}$$

and a similar equality holds with p replaced by q. By Lemma 7.21, we have

$$\|U\|_p = \lim_{n\to\infty}\left(\frac{1}{2^{2n}}\sum_{i=1}^{2^n}\sum_{j=1}^{2^n}|A_n(i,j)|^p\right)^{1/p} \leq \lim_{n\to\infty}\left(\frac{1}{2^{2n}}\sum_{i=1}^{2^n}\sum_{j=1}^{2^n}|A_n(i,j)|^q\right)^{1/q} = \|U\|_q.$$

The result follows. □

Since all of the ℓ_p-norms output the same value when U is a constant function, one should interpret Proposition 7.20 as saying that the ℓ_p-norm becomes “more sensitive” to fluctuations in U for larger values of p. In particular, the least sensitive of these norms is the ℓ_1-norm. Let us see how the cut norm compares to the ℓ_p-norms in this respect. As the next proposition and the example that follows it demonstrate, the cut norm is far less sensitive than any of the ℓ_p-norms.

Proposition 7.22. *For every kernel U, we have $\|U\|_\square \leq \|U\|_1$.*

Proof. By the definition of $\|\cdot\|_\square$ and $\|\cdot\|_1$ and the triangle inequality,

$$\|U\|_\square = \sup_{S,T\subseteq[0,1]} \left|\int_{S\times T} U(x,y)dxdy\right| \leq \sup_{S,T\subseteq[0,1]} \int_{S\times T} |U(x,y)|dxdy = \int_{[0,1]^2} |U(x,y)|dxdy = \|U\|_1.$$

$\square$

Example 7.23. Let A be an $n \times n$ symmetric matrix such that each entry is equal to -1 or 1, each with probability $1/2$, independently of one another. Let $U = U_A$, as defined in Example 7.8. Then, clearly,

$$\|U\|_1 = 1.$$

Now, let us think about the cut norm. For any two sets $S, T \subseteq [n]$ the expected value of $\sum_{i\in S}\sum_{j\in T} A(i,j)$ is zero. Intuitively speaking, that suggests that, with high probability, every pair of sets S and T satisfies $\left|\sum_{i\in S}\sum_{j\in T} A(i,j)\right| = o(n^2)$. This implies that $\|U\|_\square \to 0$ as $n \to \infty$. To make this rigorous, one can use the Chernoff Bound (Theorem B.10); see Exercise 7.5.

7.4 The Counting Lemma

In the context of extremal combinatorics, the fact that the cut norm is insensitive to local fluctuations in a matrix should be seen as a virtue rather than a flaw. Consider, for example, two random graphs G_1 and G_2 on n vertices sampled according to $G(n, 1/2)$, independently of one another. Let $W_i = W_{G_i}$ for $i \in \{1, 2\}$, as defined in Example 7.8. From the perspective of the ℓ_1-norm, the graphons W_1 and W_2 are very different. That is,

$$\|W_1 - W_2\|_1 \approx \frac{1}{2}$$

with high probability. However, from the perspective of extremal graph theory, the graphs G_1 and G_2 are virtually the same. That is, for every graph H, if n is large enough, then we expect G_1 and G_2 to contain essentially the same number of copies of the graph H. Therefore, when one is trying to judge whether two graphs are similar in structure, the ℓ_1-norm can be quite misleading. The problem is that the ℓ_1-norm is preoccupied with small local perturbations which do not describe the global structure. It cannot see the forest for the trees. In contrast, a slight modification of the argument in Example 7.23 shows that $\|W_1 - W_2\|_\square$ tends to zero as n tends to infinity. As it turns out, the cut norm is perfectly suited for determining whether two graphs are similar in terms of homomorphism densities.

NOT FINISHED.

Need to show that the cut norm is "attained." Perhaps that should go earlier?hh Define weak isomorphism and $\delta_\square(U, W)$ as well. Should define the cut metric $\delta_\square$. Need to explain how this is even better for extremal combinatorics because it recognizes that homomorphism densities are invariant under permuting vertices.

Our goal in this section is to prove the following result of Lovász and Szegedy [187] which shows that any two matrices which are close in cut norm have similar homomorphism densities.

Lemma 7.24 (The Counting Lemma)**.** *Let W_1 and W_2 be graphons. Then, for any graph H,*

$$|t(H, W_1) - t(H, W_2)| \leq |E(H)| \cdot \delta_\square(W_1, W_2).$$

Corollary 7.25. *If $(W_n)_{n=1}^{\infty}$ is a sequence of graphons and W is a graphon such that*

$$\lim_{n\to\infty} \delta_\square(W_n, W) = 0,$$

then $(W_n)_{n=1}^{\infty}$ left-converges to W.

MAYBE it would be good to mention that there exists an inverse counting lemma, and so convergence in $\delta_\square$ is equivalent to left-convergence; however, we won't need the inverse counting lemma it in these notes, so we won't go there.

7.5 The Weak Regularity Lemma

Define the "stepping" operation. OR PERHAPS it should have been defined back when we used it to prove Proposition 7.20.

Lemma 7.26 (The Weak Regularity Lemma [108])**.** *For every kernel U and $k \geq 1$ there exists a partition $\mathcal{P}$ of $[0,1]$ into at most 2^{k^2} measurable subsets such that*

$$\|U - U_{\mathcal{P}}\|_\square \leq \frac{2\|U\|_2}{k}.$$

The proof of the Weak Regularity Lemma follows somewhat of a similar strategy (in spirit) to that of the regularity lemma.

Lemma 7.27. *For every kernel U there are two measurable sets $S, T \subseteq [0,1]$ and a real number*

7.6 Proof of the Lovász–Szegedy Compactness Theorem

Need the Martingale Convergence Theorem. Well, actually, this is probably enough [184, Proposition 9.8].

7.7 Exercises

7.1 The following is a standard fact:

Fact: Let $p > 1$ and $x_1, \ldots, x_n$ be real numbers. If

$$\left(\sum_{i=1}^{n} x_i\right)^p = n^{p-1}\left(\sum_{i=1}^{n} |x_i|^p\right),$$

then $x_1 = x_2 = \cdots = x_n$.

Using the above fact, prove the following strengthening of Theorem 4.8: $t(C_4, G) > t(K_2, G)$ for every finite graph G.

7.2 Prove that the cut norm for kernels defined in Definition 7.17 is a norm.

7.3 Prove that $t(H, A_G) = t(H, G)$ for every graph H, where A_G is the adjacency matrix of G.

7.4 (a) Prove that, for every graphon W,

$$\sup_{A \subseteq [0,1]} \left| \int_A \int_A W(x,y)dxdy \right| \geq \frac{1}{2}\|W\|_\square$$

where the supremum is over all measurable subsets of $[0,1]$.

(b) Prove that, for every graphon W,

$$\sup_{\substack{S,T \subseteq [0,1] \\ S \cap T = \emptyset}} \left| \int_S \int_T W(x,y)dxdy \right| \geq \frac{1}{4}\|W\|_\square$$

where the supremum is over all pairs of disjoint measurable subsets of $[0,1]$.

(c) Prove that, for every graphon W,

$$\sup_{A \subseteq [0,1]} \left| \int_A \int_{[0,1]\setminus A} W(x,y)dxdy \right| \geq \frac{1}{6}\|W\|_\square$$

where the supremum is over all measurable subsets of $[0,1]$.

(d) Prove that, for every graphon W,

$$\sup_{\substack{S,T \subseteq [0,1] \\ \mu(S),\mu(T) \geq 1/2}} \left| \int_S \int_T W(x,y)dxdy \right| \geq \frac{1}{4}\|W\|_\square$$

where the supremum is over all pairs of measurable subsets of $[0,1]$, each of measure at least $1/2$.

(e) Find an example of a graphon W for which the inequality in part (a) is attained.

7.5 Something about applying Chernoff to make Example 7.23 rigorous.

7.6 Let $(W_n)_{n=1}^\infty$ be a sequence of graphons and let W be a $\{0,1\}$-valued graphon. Suppose that $\lim_{n\to\infty} \|W_n - W\|_\square = 0$. Prove that $\lim_{n\to\infty} \|W_n - W\|_1 = 0$ as well.

7.7 AN EXERCISE about weak* convergence and how it is not the same as convergence in the cut norm.

7.8 An exercise about defining the cut norm in terms of functions instead of sets.

7.9 An exercise about the result of the paper "strong independence of the graphcopy function".

7.10 An exercise about how the disjoint union of cliques of order 2^{-n} is not finitely forcible.

7.11 An exercise about deriving the original regularity lemma from the weak one?

7.12 Take the graphon W consisting of "dyadic cliques." Show that, for any n, there is a different graphon W' such that $t(H,W) = t(H,W')$ for all graphs H on at most n vertices.

7.8 Challenge Problems

7.1*

Appendix A

Asymptotic Notation

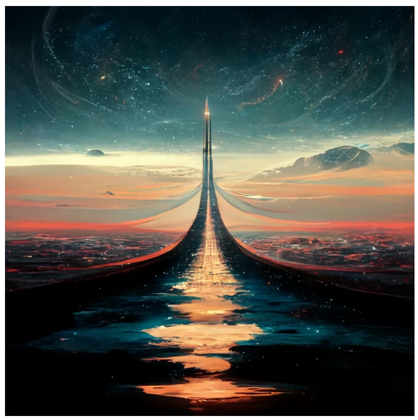


Figure A.1: "Asymptote."

Big O notation, sometimes called *Landau notation*, provides a simple shorthand for comparing the growth rates of different functions. Typically, this notation is used to describe the growth rate as a given parameter tends to infinity. Given two real-valued functions f and g, we write

- $f = O(g)$ (pronounced *f is Big O of g*) to mean that there exists constants $C > 0$ and x_0 such that $f(x) \le Cg(x)$ whenever $x \ge x_0$.
- $f = \Omega(g)$ (pronounced *f is Omega of g*) to mean that there exists constants $c > 0$ and x_0 such that $f(x) \ge cg(x)$ whenever $x \ge x_0$.
- $f = \Theta(g)$ (pronounced *f is Theta of g*) to mean that there exists constants $C > c > 0$ and x_0 such that $cg(x) \le f(x) \le Cg(x)$ whenever $x \ge x_0$.
- $f = o(g)$ or $f \ll g$ (pronounced *f is little o of g*) to mean that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.
- $f = \omega(g)$ or $f \gg g$ (pronounced *f is little omega of g*) to mean that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = \infty$.
- $f = (1 + o(1))g$ or $f \sim g$ (pronounced *f is asymptotic to g*) to mean that $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Of course, there is some redundancy here, as some of these notations can be described in terms of the others. For example, $f = \Theta(g)$ is the same as $f = O(g)$ and $f = \Omega(g)$ and $f \sim g$ is the same as $f - g = o(g)$, which is the same as $f - g = o(f)$.

Example A.1. If $f(x) = x^2 - 12x$, $g(x) = x^3$ and $h(x) = 7x - 3$, then the following are true

- $f = O(g)$,
- $f = o(g)$,
- $g = \Omega(f)$,
- $fh = \Theta(g)$,
- $f = (1 + o(1))x^2$,
- $gh \sim 7x^4$,
- $h = o(e^x)$,
- $\sin(x) = o(h)$,
- $\sin(x) + 100 = \Theta(1)$,
- $e^{-x} - 4 = \Theta\left(\frac{x^2-6}{7x^2+2x-200}\right)$,
- $f = o(x^2 \log(x))$,
- $1/f = o(1)$,
- $1/g = \Omega(e^{-x})$.

Example A.2. In this course, most of the functions that we deal with are combinatorial in nature. Thus, the variable that we are dealing with is usually not x, but some parameter of the combinatorial objects that we are considering. Here are some examples of very simple combinatorial facts written in Big O notation:

- If G is a finite graph, then $|E(G)| = O(|V(G)|^2)$.
- If G is a tree, then $|E(G)| = (1 + o(1))|V(G)|$.
- For fixed k, we have $\binom{n}{k} = \Theta(n^k)$.

Big O notation can also be used to compare two functions as the variable tends to a particular value c, as opposed to infinity. When doing so, one usually quantifies it by appending "as $x \to c$". For example, one can say that $1/(x-1)^2 = o(1/|x-1|)$ as $x \to 1$ or that $\sin(x)/x = \Theta(1)$ as $x \to 0$.

If you would like to do some more reading to solidify your understanding of asymptotics and asymptotic notation, there are many resources that you can find on the internet by doing a search for "Big O notation," or a similar keyword. Some of what you will find may be helpful and well-written, and some might just add to your confusion. To save yourself some time and effort, you might want to start by reading this objectively well-written chapter of Fleck [97]: `http://mfleck.cs.illinois.edu/building-blocks/version-1.3/big-o.pdf`.

A.1 Exercises

A.1 Determine whether each of the following is true or false. The answers are in footnotes, written upside down to help you avoid spoilers[1,2].

(a) $x^7 + x^5 = o(x^9)$.

(b) $x^{100} = O(2^x)$.

(c) $x^9 = O(x^{10}\sin(x))$.

(d) $x^2 - 9x = \Theta(x^2)$.

(e) $e^{x^2-9x} = \Theta(e^{x^2})$.

(f) $n^8 = 2^{O(\log(n))}$.

(g) $3^{\sqrt{\log(n)}} = \Omega(n)$.

(h) $7^{\log\log(n)} = \log(n)^{O(1)}$.

(i) $|E(K_n)| = (1+o(1))(n^2/2)$.

(j) $\binom{n}{5} \sim n^5$.

(k) $e^{2\pi} = \Omega(1)$.

(l) $e^{-x} + x = \Theta(1)$ as $x \to 0$.

(m) $e^{-x} + x = 1 + o(1)$ as $x \to 0$.

(n) $\frac{7n^3-4n^2+\pi n}{42n^2-22n+77} = \Theta(n)$.

(o) $n! = O(n^n)$.

(p) If v is a vertex of a graph G with n vertices, then the number of triangles in G containing v is $O(n^2)$.

(q) For every $n \geq 1$, there exists a graph G with n vertices and $\Omega(n^3 \log\log(n))$ triangles.

(r) The number of subsets of $[n]$ is $2^{O(n)}$.

(s) The number of subsets of $[n]$ is $2^{o(n)}$.

(t) The number of subsets of $[n]$ is $\Omega(2^n)$.

(u) The number of subsets of $[n]$ is $(1-o(1))2^n$.

(v) The number of subsets of $[n]$ is $\Theta(n)$.

(w) If $f(n) = \Theta(n^2)$, then $\lim_{n\to\infty} \frac{f(n)}{n^2}$ exists.

A.2 Find three functions $f, g, h : \mathbb{N} \to \mathbb{N}$ such that all four of the following conditions hold:

- $f(n) > h(n)$ for all $n \geq 1$,
- $g(n) > h(n)$ for all $n \geq 1$,
- $f \sim g$,
- $f - h \nsim g - h$.

[1] (a) T, (b) T, (c) F, (d) T, (e) F, (f) T, (g) F, (h) T, (i) T, (j) F, (k) T, (l) T, (m) T,

[2] (n) T, (o) T, (p) T, (q) F, (r) T, (s) F, (t) T, (u) T, (v) F, (w) F

A.3 Give an example of two functions $f, g : \mathbb{N} \to \mathbb{N}$ such that $f = \Theta(g)$ but there does not exist $c \in \mathbb{R}$ such that $f \sim c \cdot g$.

A.4 Let $\ell \leq k$ be a fixed positive integers and let n be an integer which is tending to infinity. Explain why the following is true:

$$\frac{\binom{37\ell^k n+92}{k}}{\binom{n-2^\ell}{k-\ell}} = \Theta(n^\ell).$$

Appendix B

Basic Discrete Probability

Figure B.1: "Discrete probability."

Everything in this appendix should be familiar to anyone who has taken a probability or statistics course before (e.g. Math 352 at UVic). In these notes, all probability spaces that we require are discrete (i.e. countable). In fact, they are all finite. Therefore, we will not need the full measure-theoretic definition of a probability space; the following will suffice.

Definition B.1. A *discrete probability space* is a pair $(\Omega, \mathbb{P})$ where X is a countable set and $\mathbb{P} : 2^\Omega \to [0, 1]$ is a function satisfying

$$\mathbb{P}(\Omega) = 1$$

and, for any $A, B \subseteq \Omega$ with $A \cap B = \emptyset$,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Definition B.2. Given a discrete probability space $(\Omega, \mathbb{P})$, the subsets E of Ω are called *events*.

Definition B.3. Given a discrete probability space $(\Omega, \mathbb{P})$, a *random variable* is a function $X : \Omega \to \mathbb{R}$.

Definition B.4. The *expectation* or *expected value* of a random variable X on a discrete probability space $(\Omega, \mathbb{P})$ is given by

$$\mathbb{E}(X) = \sum_{x \in \Omega} X(x)\mathbb{P}(\{x\}).$$

Observation B.5 (Linearity of Expectation)**.** *For any two random variables X and Y on a discrete probability space $(\Omega, \mathbb{P})$ and $a, b \in \mathbb{R}$,*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

Proof. Easily deduced from the definition of expectation. □

Definition B.6. Given a discrete probability space $(\Omega, \mathbb{P})$ and an event E, the *indicator random variable* of E is the random variable I_E such that

$$I_E(x) = \begin{cases} 1 & \text{if } x \in E, \\ 0 & \text{otherwise.} \end{cases}$$

Observation B.7. *For any event E in a discrete probability space $(\Omega, \mathbb{P})$, it holds that*

$$\mathbb{E}(I_E) = \mathbb{P}(E).$$

Proof. Easily deduced from the definition of expectation and of $\mathbb{P}$. □

The simplest tool from the toolbox of the "probabilistic method" in combinatorics, and the only one that we will use in this course, is the *First Moment Method*. This is based on the following simple principle.

Observation B.8 (First Moment Principle)**.** *If X is a random variable such that $\mathbb{E}(X) \leq \mu$, then*

$$\mathbb{P}(X \leq \mu) > 0.$$

Proof. We prove the contrapositive. Suppose that $\mathbb{P}(X \leq \mu) = 0$. That is,

$$\mathbb{P}(x \in \Omega : X(x) \leq \mu) = 0.$$

Then all non-zero terms in the sum in the definition of $\mathbb{E}(X)$ are greater than μ. Since probabilities sum to one, this tells us that $\mathbb{E}(X) > \mu$. □

The way in which we typically apply the First Moment Principle is described as follows. Suppose that we choose a certain type of mathematical object (e.g. a graph, a function or a set) randomly from a particular collection of objects. If the randomly chosen object satisfies a particular inequality *on average*, then there must be at least one element of the collection which satisfies the inequality. Let us demonstrate the method with a quick and easy application of it due to Erdős [75].

Theorem B.9 (Erdős [75])**.** *If $\mathcal{F}$ is a collection of fewer than 2^{n-1} subsets of a finite set X of cardinality n, then it is possible to colour the elements of X red and blue so that no set in $\mathcal{F}$ is monochromatic.*

Proof. We flip a coin for each element of X. If the coin flip for x is heads, we colour it red and if it is tails we colour it blue. For a given set $S \subseteq X$ of size n, the probability that every element of S is red is

$$\left(\frac{1}{2}\right)^n.$$

Likewise, the probability that every element is blue is the same. Thus, the expected number of sets in $\mathcal{F}$ whose elements all get the same colour is

$$|\mathcal{F}|\left(\left(\frac{1}{2}\right)^n + \left(\frac{1}{2}\right)^n\right) = \frac{|\mathcal{F}|}{2^{n-1}}$$

which is less than one since $|\mathcal{F}| < 2^{n-1}$. Therefore, by the First Moment Principle, there must exist a colouring in which fewer than one set in $\mathcal{F}$ is monochromatic. Since the number of monochromatic sets is an integer, it must be zero. $\square$

Let us also mention one more classical result, known as the *Chernoff Bound*. It will only be used in the analysis of some simple examples.

Theorem B.10 (The Chernoff Bound; see, e.g., [67, Theorem 1.1])**.** *Let $X_1, \ldots, X_n$ be independent random variables taking values in $[0, 1]$ and let $X = \sum_{i=1}^n x_i$. Then*

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq 2e^{-2t^2/n}.$$

B.1 Exercises

B.1 Colour each edge of the complete graph on n vertices red with probability p and blue with probability $1 - p$, independently of one another. What is the expected number of red copies of K_r? The expected number of blue copies of K_s?

B.2 Let G be a bipartite graph with n vertices. Suppose that each vertex of G is assigned a list $L(v)$ of more than $\log_2(n)$ colours. Show that there is a proper colouring of G in which, for every vertex v, the colour of v is contained in $L(v)$.

B.3 For $n \geq r \geq 1$, let $\mathcal{F} \subseteq \binom{[n]}{r}$. Prove that it is possible to colour $[n]$ with r colours such that at least $\frac{r!}{r^r}|\mathcal{F}|$ elements of $\mathcal{F}$ contain a point of every colour.

For further reading on the probabilistic method in combinatorics, see the books of Alon and Spencer [10] and Molloy and Reed [198].

Appendix C

Convexity

Figure C.1: "Convexity."

Throughout the notes, we apply many "convexity" arguments. Most of these involve "sums of squares," see Corollary C.7 below. Convexity also arises in some of the entropy-based proofs.

Definition C.1. Let $I \subseteq \mathbb{R}$ be an interval (of possibly infinite length). A function $\varphi : I \to \mathbb{R}$ is said to be *convex* if, for all $x, y \in I$ and $t \in [0,1]$,

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y). \tag{C.2}$$

Recall that a subset S of $\mathbb{R}^n$ is *convex* if, for every $x, y \in S$ and $\lambda \in [0,1]$, it holds that $\lambda x + (1-\lambda)y \in S$. This provides us with another way to think of a convex function. A function is convex if and only if the region in $\mathbb{R}^2$ above the graph of the function is a convex subset of $\mathbb{R}^2$.

Observation C.3. *A twice-differentiable function $\varphi : I \to \mathbb{R}$ is convex if and only if its second derivative is positive everywhere on I.*

In particular, all of the functions x^p for $p \geq 1$ are convex. The function $x \log(x)$ on $(0, \infty)$ is also convex. The following theorem, known as Jensen's Inequality, is a fundamental fact about convex functions which is used throughout mathematics.

Theorem C.4 (Jensen's Inequality)**.** *If φ is a convex function, then for any $z_1, \ldots, z_n$ in the domain of φ and any positive numbers $a_1, \ldots, a_n$,*

$$\varphi\left(\frac{\sum_{i=1}^n a_i z_i}{\sum_{i=1}^n a_i}\right) \leq \frac{\sum_{i=1}^n a_i \varphi(z_i)}{\sum_{i=1}^n a_i}.$$

Proof. We proceed by induction on n. The case $n = 1$ is trivial, as both sides of the inequality evaluate to $\varphi(z_i)$. Now, suppose that $n \geq 2$. Define

$$A := \sum_{i=1}^n a_i$$

$$t := \frac{a_1}{A},$$
$$x := z_1,$$

and

$$y := \sum_{i=2}^n \frac{a_i z_i}{A - a_1}.$$

Since I is an interval (which is a convex set) and $\sum_{i=2}^n \frac{a_i}{A-a_1} = 1$, we have $y \in I$. Therefore, by (C.2),

$$\begin{aligned}\varphi\left(\frac{\sum_{i=1}^n a_i z_i}{\sum_{i=1}^n a_i}\right) &= \varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y) \\ &= \frac{a_1\varphi(z_1)}{A} + \frac{(A-a_1)\varphi(y)}{A}.\end{aligned}$$

By induction,

$$\varphi(y) \leq \sum_{i=2}^n \frac{a_i \varphi(z_i)}{A - a_1}.$$

Putting the last two inequalities together completes the proof. □

As a consequence, we obtain a version of Hölder's Inequality.

Lemma C.5 (Hölder's Inequality)**.** *Let $x_1, \ldots, x_n, y_1, \ldots, y_n$ be real numbers and let $p, q > 1$ be such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q}.$$

Proof. We assume all of the x_i and y_i are positive (to avoid dividing by zero). The general case will follow easily from this. For $1 \leq i \leq n$, let

$$z_i := \frac{|x_i|}{|y_i|^{q-1}}$$

and

$$a_i := |y_i|^q.$$

The function x^p is convex, and so Jensen's Inequality tells us that

$$\left(\frac{\sum_{i=1}^n a_i z_i}{\sum_{i=1}^n a_i}\right)^p \leq \frac{\sum_{i=1}^n a_i z_i^p}{\sum_{i=1}^n a_i}$$

$$\implies \frac{\sum_{i=1}^n a_i z_i}{\sum_{i=1}^n a_i} \leq \left(\frac{\sum_{i=1}^n a_i z_i^p}{\sum_{i=1}^n a_i}\right)^{1/p}$$

$$\implies \sum_{i=1}^n a_i z_i \leq \left(\sum_{i=1}^n a_i z_i^p\right)^{1/p} \left(\sum_{i=1}^n a_i\right)^{(p-1)/p}.$$

Note that $\frac{p-1}{p} = \frac{1}{q}$, and so we can substitute that in for the last exponent in the above expression. Let's also plug in the values of z_i and a_i. We get

$$\sum_{i=1}^n |y_i|^q \left(\frac{|x_i|}{|y_i|^{q-1}}\right) \leq \left(\sum_{i=1}^n |y_i|^q \left(\frac{|x_i|}{|y_i|^{q-1}}\right)^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q}$$

$$\implies \sum_{i=1}^n |x_i y_i| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{1/p} \left(\sum_{i=1}^n |y_i|^q\right)^{1/q}$$

which is exactly what we set out to prove. □

Corollary C.6. *For any $p > 1$ and real numbers $x_1, \ldots, x_n$,*

$$\left(\sum_{i=1}^n x_i\right)^p \leq n^{p-1} \left(\sum_{i=1}^n |x_i|^p\right).$$

Proof. Let $q = \frac{p}{p-1}$ and define $y_i = 1$ for $1 \leq i \leq n$. Apply Lemma C.5. □

In particular, in the case $p = 2$, we get the following:

Corollary C.7. *For any real numbers $x_1, \ldots, x_n$,*

$$\left(\sum_{i=1}^n x_i\right)^2 \leq n \left(\sum_{i=1}^n x_i^2\right).$$

C.1 Exercises

C.1 Show that, for any sequence $a_1, \ldots, a_n$ of real numbers,

$$\left(\sum_{i=1}^n a_i\right)^2 \leq \left(\sum_{i=1}^n |a_i|^{2/3}\right) \left(\sum_{i=1}^n |a_i|^{4/3}\right).$$

The exercise in this appendix comes from the *Cauchy–Schwarz Master Class* [248], which you may want to have a look at for further reading about the inequality, its relatives, and its applications. The following link should take you to it: `http://www.ma.huji.ac.il/~ehudf/courses/Ineq09/The%20Cauchy-Schwarz%20Master%20Class%20.pdf`.

Appendix D

Further Remarks and References

Figure D.1: "My brain is still open."

D.1 Intersection, Complexity and Correlation

The topic of intersection theorems in extremal combinatorics is very well-studied, and we have only skimmed the surface here. In fact, we have purposely omitted several of the most fundamental results of this theory (e.g. Fisher's Inequality,[1] the Odd/Eventown Theorems, the Frankl–Wilson Theorems, the Two Families and Skew Two Families Theorems, the Complete Intersection Theorem, etc) because some of these topics will be covered in a new course on algebraic methods in combinatorics that will (most likely) be lectured by Natasha Morrison at UVic next year. Anyone who is interested in learning more about these topics is strongly encouraged to attend that course; however, if you won't be able to attend the course, then you may want to have a look at these notes [169, 232] or this survey [105].

The simple combinatorial (or, probabilistic, depending on your viewpoint) proof of the Erdős–Ko–Rado Theorem presented here is due to Katona [145] and is very convenient for a first exposure to intersecting families. It is short and sweet and does not require much background. However,

[1] Fisher's Inequality is covered in Math 322 at UVic [192, Theorem 6.9].

the downside of the simple combinatorial proofs of the Erdős–Ko–Rado Theorem (of which there are several; see [104]) is that they tend to fall flat on their faces when one tries to apply them in other settings, or to obtain "stability-type" extensions. A more robust "spectral" approach to intersection problems was developed primarily by Friedgut and Dinur in the early to mid-2000s and has been successful in solving a number of important problems in the area; these include the problem of determining the maximum size of a family of graphs on the same vertex set, any two of which share a triangle [72] and the maximum size of a family of k-intersecting permutations [73]. See the notes of Filmus [94] for an introduction to these methods and a survey of some of the fruits of this line of research.

One may observe that there is a rather large gap between the upper bound on the size of a system of sets of size k without a sunflower with p petals given by Theorem 1.7 and the lower bound implied by the construction in Exercise 1.19. That is, for fixed p, the lower bound grows exponentially with k while the upper bound is exponential in $k\log(k)$ (see Stirling's Approximation). The famous Sunflower Conjecture of Erdős and Rado [84] claims that the lower bound is the true behaviour, up to changing the base of the exponent; that is, for every fixed p, there is a constant $c = c(p)$ such that every collection of c^k sets of size k contains a sunflower with p petals. In November 2015, the 10th Polymath Project[2] was launched on Gil Kalai's blog [139], devoted to the Sunflower Conjecture; this led to new ideas but ended in May 2016 without any major successes. A breakthrough on the conjecture by Alweiss, Lovett, Wu and Zhang [12] has reduced the upper bound to roughly $(\log(k))^k$; see Rao [214] for a simplification of their proof.

VC-dimension is an important topic with many applications. The applications in combinatorics alone are far too numerous to list; a couple of applications which come to mind are [5, 43]. Beyond combinatorics, VC-dimension has also had a high influence in many different fields such as, for example, machine learning and computational geometry. The "VC" in VC-dimension stands for Vapnik and Chervonenkis in recognition of a paper of theirs from the late 60s; see [263] for a reproduction of an English translation of their paper. The proof of the Sauer–Shelah Theorem given here is not one of the original ones found by Sauer or Shelah; it was discovered much more recently by Alon and, independently, Aharoni and Holzman; see [15, 138].

Extra Videos

- Timothy Gowers (Cambridge) on Intersecting Families: `https://youtu.be/gesQpnQNyqw`.
- Rob Morris (IMPA) on Intersecting Families: `http://www.youtube.com/watch?v=WEFiI4YJizU&t=35m0s`
- Po-Shen Lo (Carnegie Mellon University):
  - Intersecting Families: `http://www.youtube.com/watch?v=7PtNURFvIHE&t=40m40s`
  - VC-Dimension: `http://www.youtube.com/watch?v=EiDJZx_VlmY&t=29m0s`
- Gonzalo Fiz Pontiveros lecturing at IMPA on Intersecting Families: `http://www.youtube.com/watch?v=Gzo2A7LS_kk&t=12m20s`
- Gyula Katona's lecture on Intersecting Families: `https://youtu.be/ExTkXSFb9_w`

[2]A Polymath project is an online mass collaboration among mathematicians with the aim to crack important open problems in mathematics.

- Ehud Friedgut's seminar talk on Intersecting Families (note: this contains spoilers for Challenge Problem 1.2*)
  - `https://youtu.be/DJnUHED_GVs`
  - An earlier version `https://youtu.be/-Zlsj9EyJeg`

Extra Reading

- Chapter 2 of the notes for this course at the University of Oxford [232]
- Pages 8-9 of these notes from the University of Cambridge [169]
- Chapter 3 of the notes for this course at McGill University [205]
- Lecture notes 3 and 8 for this course at EPFL [261]
- Lecture 12 for this course at Princeton University [100]

D.2 Chains, Antichains and Shadows

Despite first being published more than thirty years ago, the books *Combinatorics* by Bollobás [32] (sometimes referred to by combinatorialists simply as "the white book") and *Combinatorics of finite sets* by Anderson [14] are still excellent sources for the topics covered in Chapters 2 and 1 of these notes. See also Gerbner and Patkós [119].

The problem of counting antichains in $2^{[n]}$, which is being explored in Exercises 2.13 and Challenge Problem 2.3*, is known as *Dedekind's Problem.* The first upper bound of the form $C^{\binom{n}{\lfloor n/2\rfloor}}$ was obtained by Hansel [129]. A remarkable result of Kleitman [152] is that the "easy" lower bound in Exercise 2.13 (a) is actually tight up to a $(1+o(1))$ factor in the exponent. That is,

$$2^{\binom{n}{\lfloor n/2\rfloor}} \leq f(n) \leq 2^{(1+o(1))\binom{n}{\lfloor n/2\rfloor}}.$$

Nowadays, the result of Kleitman [151] can be proven in a short and straightforward manner by combining the results of Challenge Problems 2.1* and 2.2* (which, themselves, have short proofs) and the so-called *container method* developed in [19, 230]; see [21, 204]. Amazingly, in 1981, Korshunov [164] obtained the precise asymptotics of $f(n)$, thereby providing an asymptotic solution to Dedekind's Problem. Unfortunately, the paper is written in Russian, and seems to be less well-known than it probably deserves to be, given the greatness of the achievement.

An interesting extension of Sperner's Theorem proposed by Osthus [208] is to consider the size of the largest antichain contained within a random family in $2^{[n]}$. To be more precise, given $p \in [0,1]$ (where we allow p to depend on n), let $2^{[n],p}$ be a random subset of $2^{[n]}$ obtained by including each subset of $[n]$ with probability p, independently of one another. One could then ask: what is the distribution of the random variable which is equal to the cardinality largest antichain $\mathcal{A}$ which is fully contained in $2^{[n],p}$? A natural way to obtain a large antichain in $2^{[n],p}$ is to consider the set

$$\binom{[n]}{\lfloor n/2\rfloor} \cap 2^{[n],p}$$

which, if p is not too small, will have cardinality $(1+o(1))p\binom{n}{\lfloor n/2\rfloor}$ with high probability (by the Chernoff Bound, Theorem B.10). Balogh, Mycroft and Treglown [20] used the container method to prove that, if $p \gg 1/n$, then this construction is asymptotically optimal; moreover, the dependence between p and n is best possible.

All of the proofs involving the container method that we have alluded to above apply a "supersaturation" (c.f. Chapter 4) extension of Sperner's Theorem. Such problems ask for the minimum number of copies of a certain substructure in a combinatorial structure of a given size. In particular, the bounds used are precisely the ones described in Challenge Problems 2.1* and 2.2*, which were first proven (in a more precise form) by Kleitman [151]; see also [20, 58, 66, 204]. In particular, Kleitman [151] proved that, for any $0 \leq m \leq 2^n$, the minimum number of pairs satisfying $A \subsetneq B$ in a collection $\mathcal{F} \subseteq 2^{[n]}$ with $|\mathcal{A}| = m$ is attained by a collection consisting of sets of cardinality as close to $n/2$ as possible. The proof is an ingenious application of Hall's Theorem and estimates on binomial coefficients. The natural extension to chains of length k, conjectured by Kleitman [151], turned out to be much harder. After several partial results in [22, 58, 66], it was finally proved by Samotij [224]. Approximate quantitative extensions of Kleitman's result to other partially ordered sets are considered in [204].

There has been a great deal of interest over the years in problems related to the Littlewood–Offord Problem, as well as so called "Inverse" Littlewood–Offord Problems [127, 167, 202, 221, 255, 256]. Such results have many applications; in particular, Natasha Morrison (at UVic) has applied results on the Inverse Littlewood–Offord Problem in a joint paper on invertibility of random symmetric matrices [41].

If you think about it, you will notice that the proofs of the LYM Inequality and Local LYM Inequality used very little about the specific structure of $2^{[n]}$. As it turns out, both of these proofs are special cases of more general arguments in so called "ranked posets" (i.e. partially ordered sets which can be split into "levels" in a consistent manner). A general theorem of Kleitman [156] says that a ranked poset satisfies a natural generalization of the LYM Inequality (called the *LYM Property*) if and only if it satisfies a generalization of the Local LYM Inequality (called the *normalized matching property*) if and only if it has a *regular covering by chains*; i.e. a multiset of maximal chains which cover any two elements of a given level the same number of times. The proof that the existence of a regular covering implies the LYM Property in general is the same as the proof of the LYM Inequality given in these notes; that is, in that proof, we are using the fact that the set of all maximal chains in $2^{[n]}$ is a regular covering. Likewise, the proof that the LYM Property implies the normalized matching property in general is the same as our derivation of the Local LYM Inequality from the LYM Inequality. The final implication that the normalized matching property implies the existence of a regular covering by chains involves "blowing up" all levels of the poset to have the same size and applying Hall's Theorem to find perfect matchings between pairs of consecutive levels. A ranked poset with the normalized matching property automatically satisfies a property analogous to Sperner's Theorem, i.e. that the biggest antichain is a level of the poset, called the *Sperner Property.* The normalized matching property and Sperner Property are very well studied; see [14] and [74].

The Kruskal–Katona Theorem is generally attributed to Kruskal [166] and Katona [144], but it was also discovered independently by a number of other researchers (see [168] for references). Not only is it a very influential result in extremal combinatorics, but the general idea of the proof (i.e. "compressions") is also an important and widely applicable technique. The proof given here is strongly influenced by the one given in these notes [169].

A version of the Harris–Kleitman Inequality was proven by Harris [130] in 1960, where he applied it to obtain results in percolation theory. The formulation in Theorem 1.16 appeared later in a paper of Kleitman [154]. This result can be seen as a special case of the well-known FKG Inequality [98] or of the even more general Four Functions Theorem of Ahlswede and Daykin [1]. See, e.g., [10, Chapter 6] for more details.

Extra Videos

- Timothy Gowers (University of Cambridge) on Sperner's Theorem: `https://youtu.be/Il-NyM5OtH8`
- Po-Shen Lo (Carnegie Mellon University) on the Littlewood–Offord Problem, Sperner's Theorem and the LYM Inequality:
  - Part 1: `https://youtu.be/E0SsRZMKhlQ`
  - Part 2: `https://youtu.be/sn3PYMnAUyE`
  - Part 3: `https://youtu.be/b1frdINrD1I` (first 40 minutes)
- Jang Soo Kim (Sungkyunkwan University) on Sperner's Theorem and the LYM Inequality: `https://youtu.be/kWBZTTdMocI`
- Gyula Katona on Sperner's Theorem and the LYM Inequality: `https://youtu.be/jKk_0lf_EDk`
- Rob Morris (IMPA) on Sperner's Theorem and the LYM Inequality: `https://youtu.be/WEFiI4YJizU` (first 34 minutes)
- Gonzalo Fiz Pontiveros lecturing at IMPA on Sperner's Theorem, the LYM Inequality and the Littlewood–Offord Problem:
  - Part 1: `http://www.youtube.com/watch?v=w5sYCMg7EHw&t=30m26s`
  - Part 2: `https://youtu.be/Gzo2A7LS_kk` (first 12 minutes)
- Andrey Kupavskii's lecture at the Institute for Advanced Study on the Kruskal–Katona Theorem: `https://youtu.be/WEycq77FcCM`

Extra Reading

- Chapter 2 of the lecture notes for this course at the University of Oxford [232]
- Chapter 1 of these lecture notes from the University of Cambridge [169]
- Chapters 1, 2 and 4 for this course at McGill University [205]
- Lecture note 1 for this course at EPFL [261]
- Lecture 12 of this course at Princeton University [100]

D.3 Classical Extremal Graph Theory

As we saw in Chapter 3, we have a very good understanding of the extremal numbers of non-bipartite graphs, at least in an asymptotic sense. However, the analogous problems for bipartite graphs are much more difficult, and there are still very few bipartite graphs H for which the growth rate of $\mathrm{ex}(n, H)$ is known. In fact, even the picture for complete bipartite graphs is not understood. See Füredi and Simonovits [112] for a survey on bipartite Turán problems.

As we saw, the Kővari–Sós–Turán Theorem provides an upper bound of $O(n^{1-1/s})$ for $\mathrm{ex}(n, K_{s,t})$ with $s \leq t$. This is known to be tight up to a constant factor for some, but not all, values of s and t. In particular, it is known that there is a matching construction when $s = 2$ [110], when $s = t = 3$ [37] (see Exercise 3.20) or when $t > (s-1)!$ [9, 159]. These examples are algebraic in nature, which is a common theme in extremal problems for bipartite graphs. More recently, an alternative approach involving "random algebraic" constructions was introduced in [23, 39] and also applied in [40, 47].

As we have mentioned, Bondy and Simonovits [35] proved that $\mathrm{ex}(n, C_{2k}) = O(n^{1+1/k})$ for all $k \geq 2$. Several improvements on the constant factor have been obtained [211, 264]. The best known bound is currently

$$\mathrm{ex}(n, C_{2k}) \leq (k-1)n^{1+1/k} + 16(k-1)n$$

due to Pikhurko [211]. Erdős and Simonovits conjectured that the Bondy–Simonovits Theorem is tight up to a constant factor; in fact, they even conjecture that, for fixed k, the sequence $\mathrm{ex}(n, C_{2k})/n^{1+1/k}$ converges as $n \to \infty$ (see [112, Conjecture 4.10]). As usual, establishing the lower bounds are the harder part. In the cases $k = 2, 3$ and 5, it is known that there are constructions which are tight up to a constant factor, but such graphs have not been found for any other k; see [112, Corollary 4.9].

Recall that Corollary 3.19 implies that, for every bipartite graph H, the extremal number $\mathrm{ex}(n, H)$ is bounded away from n^2 by a polynomial factor. However, a more subtle question is whether, for every such graph H, the extremal number grows at a polynomial rate. Erdős and Simonovits (see [112]) conjecture that, for every bipartite graph H, there exists a rational $\varepsilon(H) > 0$ and $c(H) > 0$ such that

$$\frac{\mathrm{ex}(n, H)}{n^{2-\varepsilon(H)}} \to c(H)$$

as $n \to \infty$. An interesting "inverse" problem to this one, which has been asked in many different forms in, e.g. [79] and [112], is that, for every rational $r \in [1, 2]$, there is a bipartite graph H with $\mathrm{ex}(n, H) = \Theta(n^r)$. Bukh and Conlon [40] have showed that this conjecture is true if one forbids a collection $\mathcal{H}$ of graphs, as opposed to a single graph H. For more results on this conjecture, see [134, 141].

The types of extremal problems discussed in this section are also natural for hypergraphs. An *r-uniform hypergraph* is an extension of the concept of a graph to edges of cardinality r; that is, the set of edges (which are called *hyperedges*) is contained in $\binom{V}{r}$ as opposed to $\binom{V}{2}$; we touch briefly on one particular Turán-type problem for 3-uniform hypergraphs in Chapter 5. One can then ask for the maximum number of hyperedges in an r-uniform graph on n vertices which forbids a given small hypergraph. This setting turns out to be far more difficult than that of graphs. Very few exact, or even asymptotic, results are known. In particular, a notorious open problem is to determine the Turán density of the complete 3-uniform hypergraph on four vertices, which can be thought of as the natural analogue of Mantel's Theorem. For a survey, see Keevash [146].

Extra Videos

- Sang-il Oum's lectures at IBS:
  - Turán's Theorem: `https://youtu.be/YwwTFxt1o5g` (after 43 minutes)
  - Erdős–Stone Theorem `https://youtu.be/qEf8iE4JazY`
- Matt DeVos' lecture on Turán's Theorem at SFU: `https://youtu.be/LkNGEm7tXmg`
- Jaehoon Kim's lectures at KAIST:
  - Turán's Theorem and the Kővari–Sós–Turán Theorem: `https://youtu.be/WLkygOdKfK8`
  - extremal numbers of cycles: `https://youtu.be/489dxHXcw_0`
- Yufei Zhao (MIT):
  - Mantel's Theorem and Turán's Theorem: `https://youtu.be/YAo1sd4kuOQ`
  - Extremal numbers for bipartite graphs: `https://youtu.be/hDwkKrWqdZE`
- Rob Morris (IMPA) on Turán numbers for complete bipartite graphs and even cycles: `https://youtu.be/ObTxPGnGEHo`
- Jesse Geneson (SJSU) on Turán numbers of cycles:
  - Part 1: `https://youtu.be/TZJstdP64SY`
  - Part 2: `https://youtu.be/ONihgJhavcA`

Extra Reading

- The notes for lectures 7-10 of this course at Princeton University [100]
- Chapter 5 of this course at McGill University [205]
- Chapter 1 of this course at Shandong University [178]
- Lectures 1, 3, 8 and 10 of this course at the University of Oxford [48]
- Chapter 2 of the notes for this course at MIT [270]
- The notes for Lecture 10 of this course at EPFL [162]
- The notes for Lectures 7, 8, 9, 11, 12, 17 and 21 of this course at Iowa State University [172]
- Chapter 1 of the notes for this course at the University of Cambridge [259]
- Lectures 1 and 2 of this course at Tel-Aviv University [234]

D.4 Beyond or Below the Extremal Number

Stability results for Turán's Theorem or, more generally, the Erdős–Stone Theroem have been known since the work of Erdős and Simonovits from the 1960s [76, 77, 85, 240]. The argument that we provided in Chapter 4 is based on a particularly simple proof of Füredi [111]. Exact stability results for Turán's Theorem have been obtained by Korándi, Roberts and Scott [163]; the case of Mantel's Theorem goes back earlier to the work of Erdős, Győri and Simonovits [81].

Stability theorems in extremal combinatorics are not only interesting in their own right, but there has been considerable success in applying stability-type results to transform asymptotic extremal results into exact bounds. This is even true in the much more difficult area of Turán-type problems for hypergraphs. For some discussion about this method, see Keevash [146, Section 5]; a few successful applications of it can be found here [147, 179, 206, 210].

Supersaturation theorems have also become increasingly important due to their crucial role in many applications of the container method [19, 230]. Supersaturation-type results in Ramsey Theory, similar to that which is discussed in Exercise 4.7, have also been applied to obtain improved bounds on diagonal Ramsey numbers [46, 258].[3]

As Exercise 4.4 suggests, Goodman's Bound is tight for infinitely many edge densities. However, what is less clear is whether it is tight for edge densities which are not of the form $1-\frac{1}{k}$. As it turns out, it is not. The problem of determining the minimum value of $t(K_3,G)$ as a function of $t(K_2,G)$ is known as the Triangle Density Problem. Despite many impressive efforts over the decades [31, 95, 185], it remained open until 2008 when Razborov [216] finally solved it. Razborov's solution was one of the first major breakthroughs of his "flag algebra method [215]" which has led to new results on numerous old and important conjectures in extremal combinatorics, thereby becoming one of the most influential methods in the field. Essentially, the flag algebra method provides a framework for applying semi-definite programming software to search for proofs of extremal theorems. This can be seen as a vast and powerful generalization of the "sum of squares" arguments which frequently appear in these notes. The real power of the method comes from the fact that, by using a computer, we can find very complex sum of squares arguments which a human could never expect to find by hand. Natural generalizations of Razborov's result from triangles to larger cliques have since been obtained by Nikiforov [203] and Reiher [217].

Theorem 4.8 verifies a special case of a famous open problem in extremal graph theory, known as Sidorenko's Conjecture [239], which says that

$$t(H,G) \geq t(K_2,G)^{|E(H)|}$$

for every bipartite graph H and graph G. A similar conjecture (which turns out to be equivalent by a so-called "tensor power trick") was made earlier by Erdős and Simonovits [88, 241]. Sidorenko's Conjecture can be thought of as saying that, for any bipartite graph H, the best way to minimize the homomorphism density of H (which is closely tied to the number of copies of H) in a graph G with a given number of vertices and edges is to distribute those edges randomly. This conjecture is now known in a wide range of cases (see, e.g., [51, 53, 54, 131, 149, 250]), but it remains unsolved in general. The smallest graph for which the conjecture is not known to be true is $K_{5,5}\setminus C_{10}$. Many of the existing results on this conjecture are proven using clever applications of Hölder's Inequality (Lemma C.5) and entropy methods.

[3]Some Ramsey Theory is covered in Math 422 at UVic [68, Chapter 6].

Sidorenko's Conjecture is intimately linked to the notion of "quasirandomness" in graphs which was first studied by Thomason [257] and Chung, Graham and Wilson [45] in the 80s. As we saw in Theorem 4.8, the homomorphism density of C_4 is minimized by random graphs. A similar result for even cycles is proved in Theorem 4.15. A cornerstone of the theory of quasirandomness is that, if G is a graph in which $t(C_4, G)$ is "close" to $t(K_2, G)^4$, then the structure of G must be structurally similar to a random graph. Of course, this description is rather vague; i.e. what does it mean for the structure of G to resemble that of a highly unstructured object, such as a random graph? The answer is, technically, that there are many answers. For instance, one of the key features of a random graph is that, the number of edges between any two large disjoint sets A and B of vertices is roughly $|A||B|t(K_2, G)$. Another feature is that $t(H, G) \approx t(K_2, G)^{|E(H)|}$ for any small graph H. Yet another is that all but one of the eigenvalues of the adjacency matrix of G are $o(n)$; c.f. Section 4.3. As it turns out, if $t(C_4, G)$ is close to $t(K_2, G)^4$, then G must posses all of these properties, and many others.

We should also mention that the problem of understanding the interactions between homomorphism densities is one of the hallmarks of the newly emerged theory of "graph limits." This theory is closely connected with the flag algebra method discussed earlier, and has introduced a new analytic perspective to fundamental concepts in extremal combinatorics such as quasirandomness [186] and the regularity [188]; see also [55, 132]. It has also had an impact on other areas, such as property and parameter testing in computer science [189]. See the book of Lovász [184] for more background.

Extra Videos

- I have been unable to find any suitable videos.

Extra Reading

- The notes for Lecture 13 of this course at Iowa State University [172]
- Chapters 2 and 3 of the lecture notes for this course at the University of Cambridge [259].
- First few paragraphs of this blog post by Tim Gowers [125]

D.5 The Regularity Lemma

The original application of the Regularity Lemma (or, a result which is closely related to it) was in Szemerédi's proof [252] that Roth's Theorem can be extended to arbitrarily long arithmetic progressions; this was first conjectured by Erdős and Turán [90]. Szemerédi's Theorem can be seen as a strengthening of van der Waerden's Theorem [265] which is covered in Math 422 at UVic [68, Theorem 7.5].

The Triangle Removal Lemma was originally proven (in a different form) by Ruzsa and Szemerédi [222], who used it to obtain the $(6,3)$-Theorem, which solved a conjecture of Brown, Erdős and Sós [246]. More generally, let $f_3(n, \ell, k)$ be the maximum number of hyperedges in a 3-uniform hypergraph that contains no k vertices that span at least ℓ edges. A natural conjecture which extends the $(6,3)$-Theorem is that $f(n, k+3, k) = o(n^2)$ for all $k \geq 3$. This is open even in the case $k = 4$ (the case $k = 3$ is just the $(6,3)$-Theorem). A result of Conlon, Fox, Sudakov and Zhao [52] says that $f(n, 10, 5) = o(n^{3/2})$.

The Removal Lemma for general graphs was stated in [6] and [109]. Solymosi [244] realized that triangle removal can yield the strengthening of Roth's Theorem, stated in the notes as Challenge Problem 5.1*, which was originally proved in [3]. For more background on removal lemmas, see the survey of Conlon and Fox [50].

Many of the results of Chapter 5 are stated in a rather qualitative way, without much information about the precise dependencies between different parameters. For example, as stated here, Roth's Theorem says that the largest cardinality of a set $A \subseteq [n]$ without a 3-term arithmetic progression is $o(n)$; however, one could ask for better asymptotics. The approach that we have taken in these notes (i.e. applying the Regularity Lemma) is unfortunately not very good for getting good quantitative bounds. The problem is that, since the upper bound on the number of parts in the final partition is a tower of 2s of height $\varepsilon^{-O(1)}$, the upper bound that it gives for Roth's theorem is like n divided by the inverse of the tower function of n, which is usually written as $\log_*(n)$. This function grows extremely slowly. Likewise, when using the Regularity Lemma, the dependencies that one gets between the various parameters in the Removal Lemma are also dreadful.

Therefore, a big question for many years was whether the bound on the number of parts in the Regularity Lemma could be improved. Remarkably, it turns out that the tower-type dependence is, in fact, necessary. This was first proved by Gowers [123], who constructed, for every $\varepsilon > 0$, a graph G_ε which does not have a partition of the type described in Theorem 5.8 in which the number of parts is less than a tower of 2s of height $\varepsilon^{-1/16}$. This remarkable result was mentioned in Gowers' Fields Medal citation [34]. An easier argument can be found in [201]. Much sharper bounds on the height of the tower required in the Regularity Lemma are now known [101].

While these lower bound constructions preclude any improvements to the Regularity Lemma itself, it still leaves some hope that the main applications of the lemma could be improved by different means; i.e. by finding new, regularity-free, proofs. This has proved fruitful. For instance, improved bounds on the Removal Lemma were proven by Fox [99].

It is a major open problem to nail down the correct order of magnitude in the $o(n)$ in Roth's Theorem. The original proof of Roth [220] already gave $n/\log\log(n)$, which is much better than the regularity argument that we gave in Chapter 5. As of 2020, the best known bound is $n/\log^{1+c}(n)$ for some tiny constant $c > 0$, proved by Bloom and Sisask [27]. See the blog post of Kalai [140] for more of the history.

Given the huge impact that the Regularity Lemma has had on extremal graph theory, a natural question is whether it can be extended to hypergraphs and, perhaps, can give new insight into very difficult extremal problems in that area. It turns out that this is possible; however, hypergraph regularity is far from being a straightforward extension of graph regularity. There are many additional difficulties in dealing with hypergraphs and, as a result, it took many years to discover the "right" version of the Hypergraph Regularity Lemma and prove it [124, 218]. Among the many fruits of the Hypergraph Regularity Lemma is a combinatorial proof of a multidimensional version of Szemerédi's Theorem which was originally proved in [113] using ergodic theory; this connection was noticed in [245].

We should also mention that there are many different variants of the Regularity Lemma with different applications. These include the Weak Regularity Lemma of Frieze and Kannan [108] which is highly influential in graph limits and the Strong Regularity Lemma of Alon, Fischer, Krivelevich and Sudakov [7] which was influential in the area of property testing [121]. Lower bounds on the number of parts for many forms of the Regularity Lemma can be found in [49].

Extra Videos

- Yufei Zhao's lectures at MIT on the Regularity Lemma and its applications:
  - https://youtu.be/vcsxCFSLyP8
  - https://youtu.be/RD9AWDdj-Yk
  - https://youtu.be/oiKLWa_0dhs
  - https://youtu.be/3IxWLibV_tU
- Jaehoon Kim's lecture on the regularity method at KAIST: https://youtu.be/w2Y9o9zMCaw
- Fan Wei's lectures at the Institute for Advanced Study on the Regularity Lemma and its applications:
  - https://youtu.be/xGMWWQXWepQ
  - https://youtu.be/Poh8ERrZwYc
- Rob Morris' lectures at IMPA on the Regularity Lemma and its applications:
  - https://youtu.be/NHnK6byNw_o
  - https://youtu.be/VtkpPbs5edo
  - https://youtu.be/XVWhHb4vgxk
- Po-Shen Lo's lectures at CMU on the Regularity Lemma and its applications
  - https://youtu.be/Ws6IukN5YjE
  - https://youtu.be/Vvk3IJIHc4o
  - https://youtu.be/04rTnjt1eBs
  - https://youtu.be/zkka07IxUmg
  - https://youtu.be/3zp7tSRz5cA
  - https://youtu.be/lVHED8Qj9bE

Extra Reading

- Lectures 4 and 5 of this course at the University or Oxford [48]
- Chapter 2 of these notes from a course at Shandong University [178]
- Chapter 3 of the notes for this course at MIT [270]
- The notes for Lectures 23, 24, 30, 31 and 32 of this course at Iowa State University [172]
- Lectures 5-8 of this course at Tel-Aviv University [234]
- Chapter 4 of the notes for this course at the University of Cambridge [259]
- Shagnik Das' notes from FU Berlin [57]

D.6 Independent Sets, Matchings and Trees

The fact that every triangle-free graph on n vertices has an independent set of cardinality $\Omega(\sqrt{n\ln(n)})$ was first proven by Ajtai, Komlós and Szemerédi [2]. The simpler proof given in Section 6.1 was discovered a few years later later by Shearer [236]. For more than a decade, it remained an open problem to determine whether this was the correct order of magnitude. This was finally proven by Kim [150] in 1995 using a randomized construction and the "Rödl Nibble" method. A new proof was discovered by Bohman [28] by analysing the so-called "triangle-free process" where a graph is constructed by starting with n vertices and adding random edges, one by one, with the restriction that the next edge is chosen uniformly at random among all edges that would not create a triangle. By analysing this process further, Bohman and Keevash [29] and Fiz Pontiveros, Griffiths and Morris [96] have independently shown that the triangle-free process ends with a graph of independence number at most $\sqrt{2n\ln(n)}$. Combined with Shearer's bound, this narrows down the value of $R(3,k)$ to within a factor of $4+o(1)$. In [96], it is conjectured that the construction coming from the triangle-free process is asymptotically optimal. In [61], it is shown that that the expected cardinality of a random independent set in a triangle-free graph of maximum degree d is at least

$$(1+o(1))\left(\frac{\ln(d)}{d}\right)n$$

which essentially matches Shearer's bound. This provides some evidence that Shearer's bound may not be tight.

Theorem 6.9 was originally conjectured by Granville at a number theory conference in Banff in 1988 (see [4]). It was proved by Alon [4] using a very different argument to the one in Section 6.2. The algorithm in the proof presented in these notes is known as the "Kleitman–Winston algorithm" [157, 158]. Similar arguments were rediscovered by other researchers, in particular, Sapozhenko [225, 226]. This algorithm has risen to prominence due to its involvement in the development of the "container method" of Balogh, Morris and Samotij [19] and Saxton and Thomason [230]. In fact, the sets $F_S \cup C_S$ constructed in the proof of Theorem 6.9 can be thought of as "containers" for the collection of all independent sets in the graph G. For more on containers for independent sets in graphs, see the survey of Samotij [223]. If you would like to learn more about the container method, then I would also recommend the lecture notes of Morris [199]. We should mention that much stronger bounds than Theorem 6.9 are now known. In fact, the construction in Example 6.7 is, in fact, optimal; see [60, 136, 268] and the survey [269].

Theorem 6.39 was originally conjectured by Minc [196], and is sometimes still referred to as the Minc Conjecture. A few years after Brégman's original proof [36], a new proof was found by Schrijver [231]. The beautiful entropy-based argument given in Section 6.4 is due to Radhakrishnan [212]; see also [56]. The Kahn–Lovász Theorem was originally proved by Kahn and Lovász, but they never published their proof. The short proof in Exercise 6.16 was discovered by Alon and Friedland [8]. The problem of upper bounding the number of matchings of different sizes in graphs with fixed degree sequence is very interesting and well studied. As in the problem of counting independent sets discussed above, the general philosophy for maximizing the number of matchings is that taking disjoint unions of complete bipartite graphs (as in Example 6.38) is always best; see, e.g., [59, 60, 107].

Bounds on the permanents of a $\{0,1\}$-valued matrix have connections to the problem of counting "high dimensional permutations" studied in [174]. A *d-dimensional permutation* is a $[n]^{d+1}$ array of 0s and 1s such that every "axis-aligned line" contains a unique 1. A 1-dimensional permutation is simply a permutation, in the usual sense. By Stirling's approximation, the number of such objects is

$$\left((1+o(1))\frac{n}{e}\right)^n .$$

A 2-dimensional permutation is the same as a Latin square. Thus, in Challenge Problem 6.1*, one is asked to show that the number of such objects is

$$\left((1+o(1))\frac{n}{e^2}\right)^{n^2} .$$

See a pattern emerging here? Linial and Luria [174] did; they adapted a proof of Brègman's Theorem to obtain an upper bound of

$$\left((1+o(1))\frac{n}{e^d}\right)^{n^d}$$

for d-dimensional permutations in general, where d is fixed and $n \to \infty$. They conjecture that the corresponding lower bound holds as well.

Theorem 6.44, which verifies Sidorenko's Conjecture (see Section D.4) was originally proved by Sidorenko himself [238]. The beautiful entropy proof presented in Section 6.5 is due to Szegedy [250] or, perhaps, Li and Szegedy [170], whose results go far beyond the case of trees. Entropy-based arguments have since become one of the main weapons in attacking Sidorenko's Conjecture [53]. Our treatment mainly follows the simplified exposition given by Gowers [125].

Extra Videos

- Timothy Gowers' lectures at the University of Cambridge on entropy and applications:
  - Axioms of entropy: `https://youtu.be/qYmsdwQ6WLA`
  - Further properties of entropy: `https://youtu.be/8wi7EckwqSc`
  - Using entropy to count paths of length 3: `https://youtu.be/YGiXG575fNI`
  - The formula for entropy `https://youtu.be/_894wyc64JM`
  - An entropy proof of Brègman's Theorem: `https://youtu.be/tcyf2Bfc9FM`

Extra Reading

- The notes for Lecture 4 of this course at Tel-Aviv University [235]
- This blog post on entropy by Tim Gowers [125]
- The notes for Lectures 2-3 of this course at the University of Illinois [18].
- Galvin's lecture notes: *Three tutorial lectures on entropy and counting* [115].
- This paper [56] gives a nice exposition of the entropy proof of Brègman's Theorem.
- Lectures 1 and 2 for this minicourse [173]

D.7 Graph Limits

I plan to write this soon.

D.8 Sources for Exercises and Challenge Problems

Exercise 1.1 also appears in the exercises for Math 322 at UVic [192, Chapter 3, Exercise 13]. Exercises 1.2, 1.21, 1.22, 1.25 and Challenge Problem 1.3* are borrowed from the course C8.3 Combinatorics at the University of Oxford [232]. Exercises 1.3 and 1.8 are from the final exam for the Part III Combinatorics Course at Cambridge in 2022 [207]. Exercise 1.4 is from [10, Chapter 6, Exercise 2]. Exercises 1.5 and 1.29 are borrowed from Gil Kalai's blog [137]. Exercise 1.7 comes from the final exam of the Part III Combinatorics Course at Cambridge in 2016; see [207]. Exercises 1.9 and 1.12 are from [183]. Exercise 1.10 is [14, Exercise 12.1]. The result of Exercise 1.11 was first proved by Li [171]. Exercise 1.13 has connections to a beautiful (and still unsolved) conjecture of Erdős and Kleitman [82]; see [38] for a result related to it. Exercise 1.14 and Challenge Problem 1.1* are borrowed from the exercises for the Combinatorics course at the University of Cambridge [169]. Exercise 1.15 is equivalent to bounding the chromatic number of a Kneser graph from above; the matching lower bound was known as Kneser's Conjecture and was proved by Lovász [182]. Exercises 1.16, 1.18 and 1.26 are from [266, Exercises 11.2.2, 11.2.6 and 12.4.42]; the last of these is based on a result of Seymour [233]. Exercise 1.27 is inspired by [14, Theorem 6.12]. Exercise 1.19 is from [135, Chapter 6, Exercise 2]. Exercise 1.23 is borrowed from a course on Extremal Combinatorics at EPFL [261]. The idea for Exercise 1.30 is from a 2013 course on Probabilistic and Extremal Combinatorics at Utrecht [142]. Challenge Problem 1.2* seems to have been first proved by Dinur and Safra [65], and several alternative proofs now exist; see, e.g. [94, 106].

Exercises 2.1, 2.14, 2.13 and 2.27 and Challenge Problem 2.3* have been borrowed from the C8.3 Combinatorics course at the University of Oxford [232]. Exercise 2.3 is inspired by the discussion in [32, p. 19]. Exercise 2.5 was first proved, in more generality, in [62]. Exercise 2.6 is a special case of a more general result of Anderson [13]. Exercise 2.7 is from an exercise sheet from the combinatorics course at McGill University [205]. The result of Exercise 2.8 is a celebrated theorem of Bollobás [30]. Exercise 2.9 Exercise 6 in Chapter 9 of [32]. Exercises 2.10, 2.11, 2.12, 2.19, 2.20 and 2.28 are from the final exams of the Part III Combinatorics Course at Cambridge in 2013, 2009, 2022, 2018, 2019 and 2007, respectively; see [207]. Exercise 2.15 is from [266, Exercise 11.2.13]. Exercise 2.16 (a) is a special case of a more general theorem of Mirsky [197]; it is covered in Math 322 at UVic [192, Theorem 3.2]. Exercise 2.17 appears in [32, Problem 3.6]. Exercise 2.18 was borrowed from a course taught at Queen Mary University in 2011 [243]. Exercise 2.21 is from [266, Exercise 11.2.18]. Exercise 2.24 is borrowed from an Extremal Combinatorics course taught at Queen Mary, University of London in 2012 [180]. Exercise 2.23 is based on an example of Gerbner et al. [118]. Exercises 2.25 and 2.26 and Challenge Problem 2.4* are borrowed from the Combinatorics course at the University of Cambridge [169]. Challenge Problem 2.2* can be found as Problem 5 in Lecture 11 of a course on Extremal Combinatorics at EPFL [261].

Exercise 3.1 is borrowed from a course on Graph Theory at EPFL [162, Lecture 10, Exercise 1]. Exercises 3.2, 3.10 and 3.21 were borrowed from [219]. Exercises 3.3, 3.4 and 3.5 were borrowed from [183]. Exercises 3.6 and 3.24 are borrowed from a Topics in Extremal Combinatorics course at Shandong University [178]. Exercise 3.7 is borrowed from a course on Extremal Combinatorics at Queen Mary, University of London [180]. Exercise 3.8 was borrowed from [266, Exercise 1.1.26]. Exercise 3.17 is from IMPA lecture notes on Extremal and Probabilistic Combinatorics [200]. Exercise 3.18 is an easier version of Problem B13 in the Problem Set for a course on Graph Theory and Additive Combinatorics at MIT [270]. Exercise 3.20 is based on a construction of Brown [37]. Exercise 3.19 is from [64, Chapter 7]; the problem of computing $z(n,s,t)$ is known as *Zarankiewicz's Problem.* Exercise 3.25 and Challenge Problem 3.1* are borrowed from [33, Chapter IV]. Exercises 3.27, 3.28, 3.29 and 3.31 are from [266, Section 11.1]; some of these have appeared in other sources earlier, see [266] for references. Exercise 3.30 was taken from [126, Exercise 441]. Exercise 3.32 is Exercise 8.7 of [32]; it was originally proved by de Caen [63].

The last few parts of Exercise 4.2 describe a possible connection between the "Ramsey multiplicity" problem and classical Ramsey numbers that we first learned from Volec (personal communication), who heard about it from someone else. Exercise 4.6 is based on a calculation in [111, p. 68]. Exercise 4.11 asks to prove Sidorenko's Conjecture for $K_{4,4}$; this was first done by Sidorenko [239]. Exercise 4.16 and Challenge Problem 4.1* ask to verify Sidorenko's Conjecture in the special case that H is a path. This is known as the Blakley–Roy Inequality [24]. A deep explanation for fact that the case of paths with an even number of vertices is more difficult than that of paths with an odd number of vertices can be found in a result of Blekherman et al. [26] which rules out the existence of a simple "sum of squares" proof of the former. Challenge Problem 4.2* asks to verify Sidorenko's Conjecture for the hypercube; this was a breakthrough result of Hatami [131]. Exercise 4.18 is borrored from a midterm exam from an Honors Combinatorics course at the University of Chicago [16]. Exercise (c) is a result that was originally proved by Lyons [191].

Exercise 5.1 is inspired by Problem 8 on the first exercise sheet of Conlon [48]. Exercise 5.2 was borrowed from [219]. Exercises 5.8 and 5.9 and Challenge Problem 5.2* are from from the third exercise sheet of a course on Extremal Graph Theory at the University of Hamburg [194]. Exercise 5.12 was borrowed from a Problem Set for a course on Graph Theory and Additive Combinatorics at MIT [270]. Exercise 5.14 was borrowed from [266, Exercise 1.1.36]. The theorem being proved in Exercise 5.15 is sometimes called the "degree version" of the Regularity Lemma. The purpose of Exercise 5.16 is to lead the reader through a tight construction for the Komlós Tiling Theorem [160]. Exercise 5.18 is from IMPA lecture notes on Extremal and Probabilistic Combinatorics [200]. Challenge Problem 5.2* was originally proved by Szemerédi [251].

Exercise 6.1 was given as a simple illustrative example in a talk of Benny Sudakov [249]; it seems to have originated in either [175] or [80]. Exercise 6.3 is from a midterm exam from an Honors Combinatorics course at the University of Chicago [16]. Exercise 6.4 is borrowed, more or less, from Exercise 1(a) of [199]. The first three parts of Exercise 6.6 is inspired by a calcuation for the hypercube in [117]; the corresponding upper bound for the hypercube is a theorem of Korshunov and Sapozhenko [165]. The last part of Exercise 6.6 comes from [114]. Exercises 6.5 and 6.8 are inspired by the notes for Lectures 2-3 from a course on the Container Method taught at the University of Illinois [18]. Exercise 6.11 is taken from [266, Exercise 11.2.36]. The inspiration for Exercise 6.15 comes from a breakthrough result of Gilmer [120] on Frankl's Union-Closed Sets Conjecture (see [103]) and its subsequent improvements in [11, 42, 209, 229]. Exercise 6.16 is meant to lead you through an argument of Alon and Friedland [8]. Exercise 6.17 is inspired by (but only tenuously related to) Conjecture 1.1 of [69]. Exercise 6.18 is inspired by Theorem 5 of [70]. Exercise 6.22 is an instance of the so called "tensor power trick" which is useful throughout mathematics (see [254]); this particular application of the trick is inspired by [51]. Exercise 6.23 is a theorem of Conlon, Fox and Sudakov [51] which was originally proved by combining the "dependent random choice method" [102] and the tensor power trick from Exercise 6.22. The fact that it can be reproved using entropy seems to have been discovered in [170, 250]. Exercise 6.24 comes from a paper of Kopparty and Rossman [161]. Challenge Problem 6.1* is Theorem 17.3 in [176]. Challenge Problem 6.2* was proved by a paper of Blekherman and Raymond [25]. Their result verified a conjecture of Erdős and Simonovits [86], which had been proven earlier (but in a different way) by Sağlam [228].

Exercise 7.4 is Exercise 8.4 in [184]. Exercise 7.6 is Problem E1 in the Problem Set for a course on Graph Theory and Additive Combinatorics at MIT [270]

Exercise A.2 was borrowed from a midterm exam of an Honours Combinatorics course at the University of Chicago [17].

Bibliography

[1] R. Ahlswede and D. E. Daykin. An inequality for the weights of two families of sets, their unions and intersections. *Z. Wahrsch. Verw. Gebiete*, 43(3):183–185, 1978.

[2] M. Ajtai, J. Komlós, and E. Szemerédi. A note on Ramsey numbers. *J. Combin. Theory Ser. A*, 29(3):354–360, 1980.

[3] M. Ajtai and E. Szemerédi. Sets of lattice points that form no squares. *Studia Sci. Math. Hungar.*, 9:9–11 (1975), 1974.

[4] N. Alon. Independent sets in regular graphs and sum-free subsets of finite groups. *Israel J. Math.*, 73(2):247–256, 1991.

[5] N. Alon, J. Balogh, B. Bollobás, and R. Morris. The structure of almost all graphs in a hereditary property. *J. Combin. Theory Ser. B*, 101(2):85–110, 2011.

[6] N. Alon, R. A. Duke, H. Lefmann, V. Rödl, and R. Yuster. The algorithmic aspects of the regularity lemma. *J. Algorithms*, 16(1):80–109, 1994.

[7] N. Alon, D. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[8] N. Alon and S. Friedland. The maximum number of perfect matchings in graphs with a given degree sequence. *Electron. J. Combin.*, 15(1):Note 13, 2, 2008.

[9] N. Alon, L. Rónyai, and T. Szabó. Norm-graphs: variations and applications. *J. Combin. Theory Ser. B*, 76(2):280–290, 1999.

[10] N. Alon and J. H. Spencer. *The probabilistic method.* Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[11] R. Alweiss, B. Huang, and M. Sellke. Improved lower bound for the union-closed sets conjecture. arXiv:2211.11731v1, 2022.

[12] R. Alweiss, S. Lovett, K. Wu, and J. Zhang. Improved bounds for the sunflower lemma. *Ann. of Math. (2)*, 194(3):795–815, 2021.

[13] I. Anderson. On the divisors of a number. *J. London Math. Soc.*, 43:410–418, 1968.

[14] I. Anderson. *Combinatorics of finite sets.* Dover Publications, Inc., Mineola, NY, 2002. Corrected reprint of the 1989 edition.

[15] R. P. Anstee, L. Rónyai, and A. Sali. Shattering news. *Graphs Combin.*, 18(1):59–73, 2002.

[16] L. Babai. Honors Combinatorics CMSC-27410/37200 Midterm. `http://people.cs.uchicago.edu/~laci/12comb/mt.pdf`, 2012.

[17] L. Babai. Honors Combinatorics CMSC-27410/Math-28400/CMSC-37200 Second Midterm. `http://people.cs.uchicago.edu/~laci/18comb/mt2.pdf`, 2018.

[18] J. Balogh. Math 595 - Container Method. `https://faculty.math.illinois.edu/~jobal/595_spring2019.html`, Accessed on July 30, 2020.

[19] J. Balogh, R. Morris, and W. Samotij. Independent sets in hypergraphs. *J. Amer. Math. Soc.*, 28(3):669–709, 2015.

[20] J. Balogh, R. Mycroft, and A. Treglown. A random version of Sperner's theorem. *J. Combin. Theory Ser. A*, 128:104–110, 2014.

[21] J. Balogh, A. Treglown, and A. Zs. Wagner. Applications of graph containers in the Boolean lattice. *Random Structures Algorithms*, 49(4):845–872, 2016.

[22] J. Balogh and A. Zs. Wagner. Kleitman's conjecture about families of given size minimizing the number of k-chains. *Adv. Math.*, 330:229–252, 2018.

[23] P. V. M. Blagojević, B. Bukh, and R. Karasev. Turán numbers for $K_{s,t}$-free graphs: topological obstructions and algebraic constructions. *Israel J. Math.*, 197(1):199–214, 2013.

[24] G. R. Blakley and P. Roy. A Hölder type inequality for symmetric matrices with nonnegative entries. *Proc. Amer. Math. Soc.*, 16:1244–1245, 1965.

[25] G. Blekherman and A. Raymond. Proof of the Erdős-Simonovits conjecture on walks. arXiv:2009.10845v1, 2020.

[26] G. Blekherman, A. Raymond, M. Singh, and R. R. Thomas. Simple graph density inequalities with no sum of squares proofs. *Combinatorica*, 40(4):455–471, 2020.

[27] T. F. Bloom and O. Sisask. Breaking the logarithmic barrier in roth's theorem on arithmetic progressions. arXiv:2007.03528v1, 2020.

[28] T. Bohman. The triangle-free process. *Adv. Math.*, 221(5):1653–1677, 2009.

[29] T. Bohman and P. Keevash. Dynamic concentration of the triangle-free process. *Random Structures Algorithms*, 58(2):221–293, 2021.

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

[31] B. Bollobás. Relations between sets of complete subgraphs. In *Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975)*, pages 79–84, 1976.

[32] B. Bollobás. *Combinatorics*. Cambridge University Press, Cambridge, 1986. Set systems, hypergraphs, families of vectors and combinatorial probability.

[33] B. Bollobás. *Modern graph theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[34] B. Bollobás. The work of William Timothy Gowers. In *Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998)*, number Extra Vol. I, pages 109–118, 1998.

[35] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *J. Combinatorial Theory Ser. B*, 16:97–105, 1974.

[36] L. M. Brègman. Certain properties of nonnegative matrices and their permanents. *Dokl. Akad. Nauk SSSR*, 211:27–30, 1973.

[37] W. G. Brown. On graphs that do not contain a Thomsen graph. *Canad. Math. Bull.*, 9:281–285, 1966.

[38] M. Bucić, S. Letzter, B. Sudakov, and T. Tran. Minimum saturated families of sets. *Bull. Lond. Math. Soc.*, 50(4):725–732, 2018.

[39] B. Bukh. Random algebraic construction of extremal graphs. *Bull. Lond. Math. Soc.*, 47(6):939–945, 2015.

[40] B. Bukh and D. Conlon. Rational exponents in extremal graph theory. *J. Eur. Math. Soc. (JEMS)*, 20(7):1747–1757, 2018.

[41] M. Campos, L. Mattos, R. Morris, and N. Morrison. On the singularity of random symmetric matrices. *Duke Math. J.*, 170(5):881–907, 2021.

[42] Z. Chase and S. Lovett. Approximate union closed conjecture. arXiv:2211.11689v1, 2022.

[43] M. Chudnovsky, R. Kim, C.-H. Liu, P. Seymour, and S. Thomassé. Domination in tournaments. *J. Combin. Theory Ser. B*, 130:98–113, 2018.

[44] F. R. K. Chung, R. L. Graham, P. Frankl, and J. B. Shearer. Some intersection theorems for ordered sets and graphs. *J. Combin. Theory Ser. A*, 43(1):23–37, 1986.

[45] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989.

[46] D. Conlon. A new upper bound for diagonal Ramsey numbers. *Ann. of Math. (2)*, 170(2):941–960, 2009.

[47] D. Conlon. Graphs with few paths of prescribed length between any two vertices. *Bull. Lond. Math. Soc.*, 51(6):1015–1021, 2019.

[48] D. Conlon. Extremal graph theory. `http://www.its.caltech.edu/~dconlon/Extremal-course.html`, Accessed on April 21, 2020.

[49] D. Conlon and J. Fox. Bounds for graph regularity and removal lemmas. *Geom. Funct. Anal.*, 22(5):1191–1256, 2012.

[50] D. Conlon and J. Fox. Graph removal lemmas. In *Surveys in combinatorics 2013*, volume 409 of *London Math. Soc. Lecture Note Ser.*, pages 1–49. Cambridge Univ. Press, Cambridge, 2013.

[51] D. Conlon, J. Fox, and B. Sudakov. An approximate version of Sidorenko's conjecture. *Geom. Funct. Anal.*, 20(6):1354–1366, 2010.

[52] D. Conlon, J. Fox, B. Sudakov, and Y. Zhao. The regularity method for graphs with few 4-cycles. *J. Lond. Math. Soc. (2)*, 104(5):2376–2401, 2021.

[53] D. Conlon, J. H. Kim, C. Lee, and J. Lee. Some advances on Sidorenko's conjecture. *J. Lond. Math. Soc. (2)*, 98(3):593–608, 2018.

[54] D. Conlon and J. Lee. Sidorenko's conjecture for blow-ups. *Discrete Anal.*, pages Paper No. 2, 13, 2021.

[55] J. W. Cooper, D. Kráľ, and T. L. Martins. Finitely forcible graph limits are universal. *Adv. Math.*, 340:819–854, 2018.

[56] J. Cutler and A. J. Radcliffe. An entropy proof of the Kahn-Lovász theorem. *Electron. J. Combin.*, 18(1):Paper 10, 9, 2011.

[57] S. Das. Applications of the Szemerédi Regularity Lemma, FU Berlin. `http://page.mi.fu-berlin.de/shagnik/notes/regularity.pdf`, Accessed on November 2, 2020.

[58] S. Das, W. Gan, and B. Sudakov. Sperner's theorem and a problem of Erdős, Katona and Kleitman. *Combin. Probab. Comput.*, 24(4):585–608, 2015.

[59] E. Davies, M. Jenssen, and W. Perkins. A proof of the upper matching conjecture for large graphs. *J. Combin. Theory Ser. B*, 151:393–416, 2021.

[60] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. Independent sets, matchings, and occupancy fractions. *J. Lond. Math. Soc. (2)*, 96(1):47–66, 2017.

[61] E. Davies, M. Jenssen, W. Perkins, and B. Roberts. On the average size of independent sets in triangle-free graphs. *Proc. Amer. Math. Soc.*, 146(1):111–124, 2018.

[62] N. G. de Bruijn, Ca. van Ebbenhorst Tengbergen, and D. Kruyswijk. On the set of divisors of a number. *Nieuw Arch. Wiskunde (2)*, 23:191–193, 1951.

[63] D. de Caen. A note on the probabilistic approach to Turán's problem. *J. Combin. Theory Ser. B*, 34(3):340–349, 1983.

[64] R. Diestel. *Graph theory*, volume 173 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, third edition, 2005.

[65] I. Dinur and S. Safra. The importance of being biased. In *Proceedings of the Thirty-Fourth Annual ACM Symposium on Theory of Computing*, pages 33–42. ACM, New York, 2002.

[66] A. P. Dove, J. R. Griggs, R. J. Kang, and J.-S. Sereni. Supersaturation in the Boolean lattice. *Integers*, 14A:Paper No. A4, 7, 2014.

[67] D. P. Dubhashi and A. Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, Cambridge, 2009.

[68] P. Dukes. Notes for Math 422: Enumeration and Ramsey Theory, University of Victoria. `http://www.math.uvic.ca/courses/2020s/math422/a01/M422-notes-2020.pdf`, Accessed on April 7, 2020.

[69] S. Durocher, D. S. Gunderson, P. C. Li, and M. Skala. Cycle-maximal triangle-free graphs. *Discrete Math.*, 338(2):274–290, 2015.

[70] Z. Dvořák, N. Morrison, J. A. Noel, S. Norin, and L. Postle. Bounding the number of cycles in a graph in terms of its degree sequence. *European J. Combin.*, 91:Paper No. 103206, 16, 2021.

[71] G. P. Egoryčev. Решение проблемы ван- дер- Вардена для перманентов. Akad. Nauk SSSR Sibirsk. Otdel., Inst. Fiz., Krasnoyarsk, 1980. Preprint IFSO-13 M.

[72] D. Ellis, Y. Filmus, and E. Friedgut. Triangle-intersecting families of graphs. *J. Eur. Math. Soc. (JEMS)*, 14(3):841–885, 2012.

[73] D. Ellis, E. Friedgut, and H. Pilpel. Intersecting families of permutations. *J. Amer. Math. Soc.*, 24(3):649–682, 2011.

[74] K. Engel. *Sperner theory*, volume 65 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1997.

[75] P. Erdős. On a combinatorial problem. *Nordisk Mat. Tidskr.*, 11:5–10, 40, 1963.

[76] P. Erdős. Some recent results on extremal problems in graph theory. Results. In *Theory of Graphs (Internat. Sympos., Rome, 1966)*, pages 117–123 (English); pp. 124–130 (French). Gordon and Breach, New York; Dunod, Paris, 1967.

[77] P. Erdős. On some new inequalities concerning extremal properties of graphs. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 77–81. Academic Press, New York, 1968.

[78] P. Erdős. On the graph theorem of Turán. *Mat. Lapok*, 21:249–251 (1971), 1970.

[79] P. Erdős. Problems and results in combinatorial analysis and graph theory. In *Proceedings of the First Japan Conference on Graph Theory and Applications (Hakone, 1986)*, volume 72, pages 81–92, 1988.

[80] P. Erdős. Problems and results in combinatorial analysis and combinatorial number theory. In *Graph theory, combinatorics, and applications, Vol. 1 (Kalamazoo, MI, 1988)*, Wiley-Intersci. Publ., pages 397–406. Wiley, New York, 1991.

[81] P. Erdős, E. Győri, and M. Simonovits. How many edges should be deleted to make a triangle-free graph bipartite? In *Sets, graphs and numbers (Budapest, 1991)*, volume 60 of *Colloq. Math. Soc. János Bolyai*, pages 239–263. North-Holland, Amsterdam, 1992.

[82] P. Erdős and D. J. Kleitman. Extremal problems among subsets of a set. *Discrete Math.*, 8:281–294, 1974.

[83] P. Erdős, C. Ko, and R. Rado. Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford Ser. (2)*, 12:313–320, 1961.

[84] P. Erdős and R. Rado. Intersection theorems for systems of sets. *J. London Math. Soc.*, 35:85–90, 1960.

[85] P. Erdős and M. Simonovits. A limit theorem in graph theory. *Studia Sci. Math. Hungar.*, 1, 1966.

[86] P. Erdős and M. Simonovits. Compactness results in extremal graph theory. *Combinatorica*, 2(3):275–288, 1982.

[87] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2):181–192, 1983.

[88] P. Erdős and M. Simonovits. Cube-supersaturated graphs and related problems. In *Progress in graph theory (Waterloo, Ont., 1982)*, pages 203–218. Academic Press, Toronto, ON, 1984.

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

[90] P. Erdős and P. Turán. On Some Sequences of Integers. *J. London Math. Soc.*, 11(4):261–264, 1936.

[91] P. Erdős. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[92] P. Erdős. Some remarks on the theory of graphs. *Bull. Amer. Math. Soc.*, 53:292–294, 1947.

[93] D. I. Falikman. Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix. *Mat. Zametki*, 29(6):931–938, 957, 1981.

[94] Y. Filmus. Spectral methods for intersection problems: Friedgut's research program in extremal combinatorics. `http://www.cs.toronto.edu/~yuvalf/DepthPaper.pdf`, 2012, Accessed on April 10, 2020.

[95] D. C. Fisher. Lower bounds on the number of triangles in a graph. *J. Graph Theory*, 13(4):505–512, 1989.

[96] G. Fiz Pontiveros, S. Griffiths, and R. Morris. The Triangle-Free Process and the Ramsey Number $R(3, k)$. *Mem. Amer. Math. Soc.*, 263(1274):0, 2020.

[97] M. M. Fleck. Building blocks for theoretical computer science. version 1.3. `http://mfleck.cs.illinois.edu/building-blocks/version-1.3/whole-book.pdf`, Accessed on April 15, 2020.

[98] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[99] J. Fox. A new proof of the graph removal lemma. *Ann. of Math. (2)*, 174(1):561–579, 2011.

[100] J. Fox. MAT 307: Combinatorics. `http://math.mit.edu/~fox/MAT307`, Accessed on April 14, 2020.

[101] J. Fox and L. M. Lovász. A tight lower bound for Szemerédi's regularity lemma. *Combinatorica*, 37(5):911–951, 2017.

[102] J. Fox and B. Sudakov. Dependent random choice. *Random Structures Algorithms*, 38(1-2):68–99, 2011.

[103] P. Frankl. Extremal set systems. In *Handbook of combinatorics, Vol. 1, 2*, pages 1293–1329. Elsevier Sci. B. V., Amsterdam, 1995.

[104] P. Frankl and R. L. Graham. Old and new proofs of the Erdős-Ko-Rado theorem. *Sichuan Daxue Xuebao*, 26(Special Issue):112–122, 1989.

[105] P. Frankl and N. Tokushige. Invitation to intersection problems for finite sets. *J. Combin. Theory Ser. A*, 144:157–211, 2016.

[106] E. Friedgut. Proof of an intersection theorem via Fourier analysis. In *European Congress of Mathematics*, pages 117–120. Eur. Math. Soc., Zürich, 2005.

[107] S. Friedland, E. Krop, and K. Markström. On the number of matchings in regular graphs. *Electron. J. Combin.*, 15(1):Research Paper 110, 28, 2008.

[108] A. Frieze and R. Kannan. Quick approximation to matrices and applications. *Combinatorica*, 19(2):175–220, 1999.

[109] Z. Füredi. Extremal hypergraphs and combinatorial geometry. In *Proceedings of the International Congress of Mathematicians, Vol. 1, 2 (Zürich, 1994)*, pages 1343–1352. Birkhäuser, Basel, 1995.

[110] Z. Füredi. New asymptotics for bipartite Turán numbers. *J. Combin. Theory Ser. A*, 75(1):141–144, 1996.

[111] Z. Füredi. A proof of the stability of extremal graphs, Simonovits' stability from Szemerédi's regularity. *J. Combin. Theory Ser. B*, 115:66–71, 2015.

[112] Z. Füredi and M. Simonovits. The history of degenerate (bipartite) extremal graph problems. In *Erdös centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 169–264. János Bolyai Math. Soc., Budapest, 2013.

[113] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. Analyse Math.*, 34:275–291 (1979), 1978.

[114] D. Galvin. On homomorphisms from the Hamming cube to $\mathbb{Z}$. *Israel J. Math.*, 138:189–213, 2003.

[115] D. Galvin. Three tutorial lectures on entropy and counting. arXiv:1406.7872v1, 2014.

[116] D. Galvin. Counting colorings of a regular graph. *Graphs Combin.*, 31(3):629–638, 2015.

[117] D. Galvin. Independent sets in the discrete hypercube. arXiv:1901.01991v1, 2019.

[118] D. Gerbner, B. Keszegh, N. Lemons, C. Palmer, D. Pálvölgyi, and B. Patkós. Saturating Sperner families. *Graphs Combin.*, 29(5):1355–1364, 2013.

[119] D. Gerbner and B. Patkós. *Extremal finite set theory.* Discrete Mathematics and its Applications (Boca Raton). CRC Press, Boca Raton, FL, 2019.

[120] J. Gilmer. A constant lower bound for the union-closed sets conjecture. arXiv:2211.09055v1, 2022.

[121] O. Goldreich, S. Goldwasser, and D. Ron. Property testing and its connection to learning and approximation. *J. ACM*, 45(4):653–750, 1998.

[122] A. W. Goodman. On sets of acquaintances and strangers at any party. *Amer. Math. Monthly*, 66:778–783, 1959.

[123] W. T. Gowers. Lower bounds of tower type for Szemerédi's uniformity lemma. *Geom. Funct. Anal.*, 7(2):322–337, 1997.

[124] W. T. Gowers. Hypergraph regularity and the multidimensional Szemerédi theorem. *Ann. of Math. (2)*, 166(3):897–946, 2007.

[125] W. T. Gowers. Entropy and Sidorenko's conjecture – after Szegedy, *Gowers's weblog.* `https://gowers.wordpress.com/2015/11/18/entropy-and-sidorenkos-conjecture-after-szegedy/`, 2015, Accessed on April 28, 2020.

[126] D. Gunderson. Graph Theory Notes, University of Manitoba. `https://mspace.lib.umanitoba.ca/handle/1993/37010?fbclid=IwAR09o_TQvyPKAafXKHeyK2xpX-bfOGxZcodZaYJABU8gfqOpPci6KECICrI`, Accessed on December 21, 2022.

[127] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Period. Math. Hungar.*, 8(3-4):197–211, 1977.

[128] T. S. Han. Nonnegative entropy measures of multivariate symmetric correlations. *Information and Control*, 36(2):133–156, 1978.

[129] G. Hansel. Sur le nombre des fonctions booléennes monotones de n variables. *C. R. Acad. Sci. Paris Sér. A-B*, 262:A1088–A1090, 1966.

[130] T. E. Harris. A lower bound for the critical probability in a certain percolation process. *Proc. Cambridge Philos. Soc.*, 56:13–20, 1960.

[131] H. Hatami. Graph norms and Sidorenko's conjecture. *Israel J. Math.*, 175:125–150, 2010.

[132] H. Hatami and S. Norine. Undecidability of linear inequalities in graph homomorphism densities. *J. Amer. Math. Soc.*, 24(2):547–565, 2011.

[133] P. Hell and J. Nešetřil. *Graphs and homomorphisms*, volume 28 of *Oxford Lecture Series in Mathematics and its Applications.* Oxford University Press, Oxford, 2004.

[134] T. Jiang, J. Ma, and L. Yepremyan. On Turán exponents of bipartite graphs. *Combin. Probab. Comput.*, 31(2):333–344, 2022.

[135] S. Jukna. *Extremal combinatorics*. Texts in Theoretical Computer Science. An EATCS Series. Springer, Heidelberg, second edition, 2011. With applications in computer science.

[136] J. Kahn. Entropy, independent sets and antichains: a new approach to Dedekind's problem. *Proc. Amer. Math. Soc.*, 130(2):371–378, 2002.

[137] G. Kalai. Extremal Combinatorics I: Extremal Problems on Set Systems, *Combinatorics and more*. `https://gilkalai.wordpress.com/2008/05/01/extremal-combinatorics-i/`, 2008, Accessed on April 10, 2020.

[138] G. Kalai. Extremal Combinatorics III: Some Basic Theorems, *Combinatorics and more*. `https://gilkalai.wordpress.com/2008/09/28/extremal-combinatorics-iii-some-basic-theorems/`, 2008, Accessed on April 10, 2020.

[139] G. Kalai. Polymath10: The Erdős Rado Delta System Conjecture, *Combinatorics and more*. `https://gilkalai.wordpress.com/2015/11/03/polymath10-the-erdos-rado-delta-system-conjecture/`, 2015, Accessed on April 10, 2020.

[140] G. Kalai. To cheer you up in difficult times 7: Bloom and Sisask just broke the logarithm barrier for Roth's theorem!, *Combinatorics and more*. `https://gilkalai.wordpress.com/2020/07/08/to-cheer-you-up-in-difficult-times-7-bloom-and-sisask-just-broke-the-logarithm-barrier-f`
2020, Accessed on July 30, 2020.

[141] D. Y. Kang, J. Kim, and H. Liu. On the rational Turán exponents conjecture. *J. Combin. Theory Ser. B*, 148:149–172, 2021.

[142] R. J. Kang and T. Müller. Probabilistic and extremal combinatorics. `http://www.math.rug.nl/~tobias/PEC/PEC2013.html`, Accessed on April 14, 2020.

[143] G. Katona. On a conjecture of Erdős and a stronger form of Sperner's theorem. *Studia Sci. Math. Hungar.*, 1:59–63, 1966.

[144] G. Katona. A theorem of finite sets. In *Theory of graphs (Proc. Colloq., Tihany, 1966)*, pages 187–207, 1968.

[145] G. O. H. Katona. A simple proof of the Erdős-Chao Ko-Rado theorem. *J. Combinatorial Theory Ser. B*, 13:183–184, 1972.

[146] P. Keevash. Hypergraph Turán problems. In *Surveys in combinatorics 2011*, volume 392 of *London Math. Soc. Lecture Note Ser.*, pages 83–139. Cambridge Univ. Press, Cambridge, 2011.

[147] P. Keevash and B. Sudakov. The Turán number of the Fano plane. *Combinatorica*, 25(5):561–574, 2005.

[148] T. Kővari, V. T. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloq. Math.*, 3:50–57, 1954.

[149] J. H. Kim, C. Lee, and J. Lee. Two approaches to Sidorenko's conjecture. *Trans. Amer. Math. Soc.*, 368(7):5057–5074, 2016.

[150] Jeong Han Kim. The Ramsey number $R(3,t)$ has order of magnitude $t^2/\log t$. *Random Structures Algorithms*, 7(3):173–207, 1995.

[151] D. Kleitman. A conjecture of Erdős-Katona on commensurable pairs among subsets of an n-set. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 215–218. Academic Press, New York, 1968.

[152] D. Kleitman. On Dedekind's problem: The number of monotone Boolean functions. *Proc. Amer. Math. Soc.*, 21:677–682, 1969.

[153] D. J. Kleitman. On a lemma of Littlewood and Offord on the distribution of certain sums. *Math. Z.*, 90:251–259, 1965.

[154] D. J. Kleitman. Families of non-disjoint subsets. *J. Combinatorial Theory*, 1:153–155, 1966.

[155] D. J. Kleitman. On a lemma of Littlewood and Offord on the distributions of linear combinations of vectors. *Advances in Math.*, 5:155–157 (1970), 1970.

[156] D. J. Kleitman. On an extremal property of antichains in partial orders. The lym property and some of its implications and applications. In *Combinatorics (Proc. NATO Advanced Study Inst., Breukelen, 1974), Part 2: Graph theory; foundations, partitions and combinatorial geometry*, pages 77–90. Math. Centre Tracts, No. 56, 1974.

[157] D. J. Kleitman and K. J. Winston. The asymptotic number of lattices. *Ann. Discrete Math.*, 6:243–249, 1980.

[158] D. J. Kleitman and K. J. Winston. On the number of graphs without 4-cycles. *Discrete Math.*, 41(2):167–172, 1982.

[159] J. Kollár, L. Rónyai, and T. Szabó. Norm-graphs and bipartite Turán numbers. *Combinatorica*, 16, 1996.

[160] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[161] S. Kopparty and B. Rossman. The homomorphism domination exponent. *European J. Combin.*, 32(7):1097–1114, 2011.

[162] D. Korándi. Graph Theory 2018. `https://www.epfl.ch/labs/dcg/page-18464-en-html/page-154211-en-html/`, Accessed on April 6, 2020.

[163] D. Korándi, A. Roberts, and A. Scott. Exact stability for Turán's theorem. *Adv. Comb.*, pages Paper No. 9, 17, 2021.

[164] A. D. Korshunov. The number of monotone Boolean functions. *Problemy Kibernet.*, (38):5–108, 272, 1981.

[165] A. D. Korshunov and A. A. Sapozhenko. The number of binary codes with distance 2. *Problemy Kibernet.*, (40):111–130, 1983.

[166] J. B. Kruskal. The number of simplices in a complex. In *Mathematical optimization techniques*, pages 251–278. Univ. of California Press, Berkeley, Calif., 1963.

[167] M. Kwan and L. Sauermann. An algebraic inverse theorem for the quadratic Littlewood-Offord problem, and an application to Ramsey graphs. *Discrete Anal.*, pages Paper No. 12, 34, 2020.

[168] D. V. Le and T. Römer. A Kruskal-Katona type result and applications. *Discrete Math.*, 343(5):111801, 12, 2020.

[169] I. Leader. Combinatorics, 2010, University of Cambridge. Notes scribed by Gareth Taylor. `https://tartarus.org/gareth/maths/notes/iii/Combinatorics_2010.pdf`, Accessed on April 6, 2020.

[170] J. L. X. Li and B. Szegedy. On the logarithimic calculus and sidorenko's conjecture. arXiv:1107.1153v1, 2011.

[171] S. Y. R. Li. An extremal problem among subsets of a set. *J. Combinatorial Theory Ser. A*, 23(3):341–343, 1977.

[172] B. Lidický. Math 608 - Extremal Graph Theory, Iowa State University. `http://lidicky.name/oldteaching/18.608/index.html`, Accessed on July 30, 2020.

[173] N. Linial. Minicourse on Geometry and Combinatorics at the Graphs and Groups Workshop. `http://people.maths.ox.ac.uk/drutu/workshopGG/`, Accessed on July 30, 2020.

[174] N. Linial and Z. Luria. An upper bound on the number of high-dimensional permutations. *Combinatorica*, 34(4):471–486, 2014.

[175] N. Linial and Y. Rabinovich. Local and global clique numbers. *J. Combin. Theory Ser. B*, 61(1):5–15, 1994.

[176] J. H. van Lint and R. M. Wilson. *A course in combinatorics*. Cambridge University Press, Cambridge, second edition, 2001.

[177] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):277–286, 1943.

[178] H. Liu. Topics in extremal combinatorics, 2019, Shandong University. `http://homepages.warwick.ac.uk/staff/H.Liu.9/ExCombNote-Aug09.pdf`, Accessed on April 9, 2020.

[179] H. Liu, O. Pikhurko, and K. Staden. The exact minimum number of triangles in graphs with given order and size. *Forum Math. Pi*, 8:e8, 144, 2020.

[180] E. Long. Extremal Combinatorics, 2012, Queen Mary, University of London. `http://web.mat.bham.ac.uk/E.Long/Ext_course/ext_comb.html`, Accessed on April 9, 2020.

[181] L. H. Loomis and H. Whitney. An inequality related to the isoperimetric inequality. *Bull. Amer. Math. Soc*, 55:961–962, 1949.

[182] L. Lovász. Kneser's conjecture, chromatic number, and homotopy. *J. Combin. Theory Ser. A*, 25(3):319–324, 1978.

[183] L. Lovász. *Combinatorial problems and exercises*. AMS Chelsea Publishing, Providence, RI, second edition, 2007.

[184] L. Lovász. *Large networks and graph limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2012.

[185] L. Lovász and M. Simonovits. On the number of complete subgraphs of a graph. II. In *Studies in pure mathematics*, pages 459–495. Birkhäuser, Basel, 1983.

[186] L. Lovász and V. T. Sós. Generalized quasirandom graphs. *J. Combin. Theory Ser. B*, 98(1):146–163, 2008.

[187] L. Lovász and B. Szegedy. Limits of dense graph sequences. *J. Combin. Theory Ser. B*, 96(6):933–957, 2006.

[188] L. Lovász and B. Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17(1):252–270, 2007.

[189] L. Lovász and B. Szegedy. Testing properties of graphs and functions. *Israel J. Math.*, 178:113–156, 2010.

[190] D. Lubell. A short proof of Sperner's lemma. *J. Combinatorial Theory*, 1:299, 1966.

[191] R. Lyons. Asymptotic enumeration of spanning trees. *Combin. Probab. Comput.*, 14(4):491–522, 2005.

[192] G. MacGillivray and P. Dukes. Notes for Math 322: Intermediate Combinatorics, University of Victoria. `http://www.math.uvic.ca/faculty/jing/322notes.pdf`, Accessed on April 7, 2020.

[193] W. Mantel. Problem 28. *Wiskundige Opgaven*, 10:60–61, 1907.

[194] S. Messuti and M. Schacht. Extremale Graphentheorie. `https://www.math.uni-hamburg.de/home/schacht/lehre/SS12/VL_ExGT.html`, Accessed on April 23, 2020.

[195] L. D. Mešalkin. A generalization of Sperner's theorem on the number of subsets of a finite set. *Teor. Verojatnost. i Primenen*, 8:219–220, 1963.

[196] H. Minc. Upper bounds for permanents of (0, 1)-matrices. *Bull. Amer. Math. Soc.*, 69:789–791, 1963.

[197] L. Mirsky. A dual of Dilworth's decomposition theorem. *Amer. Math. Monthly*, 78:876–877, 1971.

[198] M. Molloy and B. Reed. *Graph colouring and the probabilistic method*, volume 23 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 2002.

[199] R. Morris. The method of hypergraph containers. `https://www.ime.usp.br/~spschool2016/wp-content/uploads/2016/07/Morris.pdf`, Accessed on April 27, 2020.

[200] R. Morris and R. Imbuzeiro Oliveira. *Extremal and probabilistic combinatorics*. Publicações Matemáticas do IMPA. [IMPA Mathematical Publications]. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2011. 28o Colóquio Brasileiro de Matemática. [28th Brazilian Mathematics Colloquium].

[201] G. Moshkovitz and A. Shapira. A short proof of Gowers' lower bound for the regularity lemma. *Combinatorica*, 36(2):187–194, 2016.

[202] H. H. Nguyen and V. H. Vu. Small ball probability, inverse theorems, and applications. In *Erdös centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 409–463. János Bolyai Math. Soc., Budapest, 2013.

[203] V. Nikiforov. The number of cliques in graphs of given order and size. *Trans. Amer. Math. Soc.*, 363(3):1599–1618, 2011.

[204] J. A. Noel, A. Scott, and B. Sudakov. Supersaturation in posets and applications involving the container method. *J. Combin. Theory Ser. A*, 154:247–284, 2018.

[205] S. Norin. Math 550 - Combinatorics, 2018, McGill University. Notes scribed by Eric P. Hanson. `http://www.math.mcgill.ca/~snorin/math550W18.html`, Accessed on April 6, 2020.

[206] S. Norin and L. Yepremyan. Turán number of generalized triangles. *J. Combin. Theory Ser. A*, 146:312–343, 2017.

[207] University of Cambridge. Part III of the Mathematical Tripos Examination Papers. `https://www.maths.cam.ac.uk/postgrad/part-iii/pastpapers/2016/index.html`, Accessed on October 27, 2022.

[208] D. Osthus. Maximum antichains in random subsets of a finite set. *J. Combin. Theory Ser. A*, 90(2):336–346, 2000.

[209] L. Pebody. Extension of a method of Gilmer. arXiv:2211.13139v1, 2022.

[210] O. Pikhurko. An exact Turán result for the generalized triangle. *Combinatorica*, 28(2):187–208, 2008.

[211] O. Pikhurko. A note on the Turán function of even cycles. *Proc. Amer. Math. Soc.*, 140(11):3687–3692, 2012.

[212] J. Radhakrishnan. An entropy proof of Bregman's theorem. *J. Combin. Theory Ser. A*, 77(1):161–164, 1997.

[213] F. P. Ramsey. On a Problem of Formal Logic. *Proc. London Math. Soc. (2)*, 30(4):264–286, 1929.

[214] A. Rao. Coding for sunflowers. *Discrete Anal.*, pages Paper No. 2, 8, 2020.

[215] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[216] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[217] C. Reiher. The clique density theorem. *Ann. of Math. (2)*, 184(3):683–707, 2016.

[218] V. Rödl and J. Skokan. Regularity lemma for k-uniform hypergraphs. *Random Structures Algorithms*, 25(1):1–42, 2004.

[219] J. Rollins. Extremal Combinatorics Problem Database. Karlsruhe Institute of Technology.

[220] K. F. Roth. On certain sets of integers. *J. London Math. Soc.*, 28:104–109, 1953.

[221] M. Rudelson and R. Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[222] I. Z. Ruzsa and E. Szemerédi. Triple systems with no six points carrying three triangles. In *Combinatorics (Proc. Fifth Hungarian Colloq., Keszthely, 1976), Vol. II*, volume 18 of *Colloq. Math. Soc. János Bolyai*, pages 939–945. North-Holland, Amsterdam-New York, 1978.

[223] W. Samotij. Counting independent sets in graphs. *European J. Combin.*, 48:5–18, 2015.

[224] W. Samotij. Subsets of posets minimising the number of chains. *Trans. Amer. Math. Soc.*, 371(10):7259–7274, 2019.

[225] A. A. Sapozhenko. On the number of independent sets in extenders. *Diskret. Mat.*, 13(1):56–62, 2001.

[226] A. A. Sapozhenko. The Cameron-Erdős conjecture. *Dokl. Akad. Nauk*, 393(6):749–752, 2003.

[227] N. Sauer. On the density of families of sets. *J. Combinatorial Theory Ser. A*, 13:145–147, 1972.

[228] M. Sağlam. Near log-convexity of measured heat in (discrete) time and consequences. In *59th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2018*, pages 967–978. IEEE Computer Soc., Los Alamitos, CA, 2018.

[229] W. Sawin. An improved lower bound for the union-closed set conjecture. arXiv:2211.11504v1, 2022.

[230] D. Saxton and A. Thomason. Hypergraph containers. *Invent. Math.*, 201(3):925–992, 2015.

[231] A. Schrijver. A short proof of Minc's conjecture. *J. Combinatorial Theory Ser. A*, 25(1):80–83, 1978.

[232] A. Scott. C8.3 Combinatorics, 2020, University of Oxford. Modified by J. Long. `https://courses.maths.ox.ac.uk/node/42882/materials`, Accessed on April 6, 2020.

[233] P. D. Seymour. On incomparable collections of sets. *Mathematika*, 20:208–209, 1973.

[234] A. Shapira. Extremal Graph Theory. `https://www.cs.tau.ac.il/~asafico/ext-graph-theory/notes.pdf`, Accessed on July 30, 2020.

[235] A. Shapira. Topics in Extremal Combinatorics. `https://www.cs.tau.ac.il/~asafico/topics-15/topics.pdf`, Accessed on July 30, 2020.

[236] J. B. Shearer. A note on the independence number of triangle-free graphs. *Discrete Math.*, 46(1):83–87, 1983.

[237] S. Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific J. Math.*, 41:247–261, 1972.

[238] A. Sidorenko. A correlation inequality for bipartite graphs. *Graphs Combin.*, 9(2):201–204, 1993.

[239] A. F. Sidorenko. Inequalities for functionals generated by bipartite graphs. *Diskret. Mat.*, 3(3):50–65, 1991.

[240] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 279–319. Academic Press, New York, 1968.

[241] M. Simonovits. Extremal graph problems, degenerate extremal problems, and supersaturated graphs. In *Progress in graph theory (Waterloo, Ont., 1982)*, pages 419–437. Academic Press, Toronto, ON, 1984.

[242] M. Simonovits. Paul Erdős' influence on extremal graph theory. In *The mathematics of Paul Erdős, II*, volume 14 of *Algorithms Combin.*, pages 148–192. Springer, Berlin, 1997.

[243] O. Sisask. MTH711U: Extremal Combinatorics, 2011, Queen Mary, University of London. `https://people.kth.se/~sisask/MTH711U/`, Accessed on August, 2022.

[244] J. Solymosi. Note on a generalization of Roth's theorem. In *Discrete and computational geometry*, volume 25 of *Algorithms Combin.*, pages 825–827. Springer, Berlin, 2003.

[245] J. Solymosi. A note on a question of Erdős and Graham. *Combin. Probab. Comput.*, 13(2):263–267, 2004.

[246] V. T. Sós, P. Erdős, and W. G. Brown. On the existence of triangulated spheres in 3-graphs, and related problems. *Period. Math. Hungar.*, 3(3-4):221–228, 1973.

[247] E. Sperner. Ein Satz über Untermengen einer endlichen Menge. *Math. Z.*, 27(1):544–548, 1928.

[248] J. M. Steele. *The Cauchy-Schwarz master class*. MAA Problem Books Series. Mathematical Association of America, Washington, DC; Cambridge University Press, Cambridge, 2004. An introduction to the art of mathematical inequalities.

[249] B. Sudakov. Large independent sets from local considerations. Probabilistic Combinatorics Online 2020 at MIPT, `http://pco20.combgeo.org/`, Viewed online on September 23, 2020.

[250] B. Szegedy. An information theoretic approach to Sidorenko's conjecture. arXiv:1406.6738v3, 2015.

[251] E. Szemerédi. On graphs containing no complete subgraph with 4 vertices. *Mat. Lapok*, 23, 1972.

[252] E. Szemerédi. On sets of integers containing no k elements in arithmetic progression. *Acta Arith.*, 27:199–245, 1975.

[253] E. Szemerédi. Regular partitions of graphs. In *Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976)*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.

[254] T. Tao. Tricks Wiki article: the tensor power trick. `https://terrytao.wordpress.com/2008/08/25/tricks-wiki-article-the-tensor-product-trick/`, 2008, Accessed on October 9, 2020.

[255] T. Tao and V. Vu. The Littlewood-Offord problem in high dimensions and a conjecture of Frankl and Füredi. *Combinatorica*, 32(3):363–372, 2012.

[256] T. Tao and V. H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.

[257] A. Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

[258] A. Thomason. An upper bound for some Ramsey numbers. *J. Graph Theory*, 12(4):509–517, 1988.

[259] A. Thomason. Extremal Graph Theory, 2017, University of Cambridge. Notes scribed by Dexter Chua. `https://dec41.user.srcf.net/notes/III_M/extremal_graph_theory_thm_proof.pdf`, Accessed on July 30, 2020.

[260] C. Thomassen. On the chromatic number of triangle-free graphs of large minimum degree. *Combinatorica*, 22(4):591–596, 2002.

[261] I. Tomon. Extremal Combinatorics 2018. `https://www.epfl.ch/labs/dcg/page-18464-en-html/page-154228-en-html/`, Accessed on April 6, 2020.

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

[263] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of complexity*, pages 11–30. Springer, Cham. Reprint of Theor. Probability Appl. **16** (1971), 264–280.

[264] J. Verstraëte. On arithmetic progressions of cycle lengths in graphs. *Combin. Probab. Comput.*, 9(4):369–373, 2000.

[265] B. L. van der Waerden. Beweis einer baudetschen vermutung. *Nieuw. Arch. Wisk.*, 15:212–216, 1927.

[266] D. B. West. *Combinatorial Mathematics.* Cambridge University Press, Cambridge, 2020.

[267] K. Yamamoto. Logarithmic order of free distributive lattice. *J. Math. Soc. Japan*, 6:343–353, 1954.

[268] Y. Zhao. The number of independent sets in a regular graph. *Combin. Probab. Comput.*, 19(2):315–320, 2010.

[269] Y. Zhao. Extremal regular graphs: independent sets and graph homomorphisms. *Amer. Math. Monthly*, 124(9):827–843, 2017.

[270] Y. Zhao. Graph theory and additive combinatorics. `https://ocw.mit.edu/courses/mathematics/18-217-graph-theory-and-additive-combinatorics-fall-2019/index.htm`, Accessed on May 14, 2020.

Figure 1: “Erdős joyfully reading The BOOK.”