

# Lecture notes for 21-301: Combinatorics

Michael Tait

Fall 2017

Carnegie Mellon University



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Enumeration</b>	<b>3</b>
2.1	Sequences and the Multiplication Principle . . . . .	3
2.2	Permutations and Combinations . . . . .	5
2.3	Bijections and Double Counting . . . . .	7
2.4	Estimation . . . . .	12
2.5	Inclusion-Exclusion . . . . .	16
2.6	Generating Functions . . . . .	20
2.7	Formal Power Series . . . . .	23
2.8	Generating Functions Redux . . . . .	25
2.8.1	Change making . . . . .	27
2.8.2	Compositions . . . . .	28
2.8.3	Counting Subsets . . . . .	29
2.8.4	Counting Strings . . . . .	30
<b>3</b>	<b>The Probabilistic Method</b>	<b>33</b>
3.1	Preliminaries . . . . .	33
3.2	The first moment method . . . . .	35

3.2.1	Linearity of expectation . . . . .	35
3.3	Alterations . . . . .	38
3.4	Markov and Chebyshev . . . . .	40
3.5	Chernoff Bound . . . . .	46
3.6	Lovász Local Lemma . . . . .	48
<b>4</b>	<b>Extremal Graph Theory</b>	<b>55</b>
4.1	Turán’s Theorem . . . . .	56
4.2	Projective planes . . . . .	62
4.3	Sidon sets . . . . .	65
4.4	Constructing $C_4$ -free graphs . . . . .	66
4.5	Ramsey numbers . . . . .	68
4.6	Combinatorial Number Theory . . . . .	71
4.7	Erdős-Ko-Rado Theorem . . . . .	76
<b>5</b>	<b>Spectral graph theory</b>	<b>79</b>
5.1	Linear Algebra Preliminaries . . . . .	80
5.2	The adjacency matrix . . . . .	83
5.3	Short proofs of old results using spectral graph theory . . . . .	86
5.4	The Graham-Pollak Theorem . . . . .	91
5.5	The Expander-Mixing Lemma . . . . .	94
5.6	The Hoffman-Singleton Theorem . . . . .	96

# 1

## Introduction

*The only way to learn mathematics is to do mathematics.*

– Paul Halmos

These are lecture notes for Math 21-301: Combinatorics, taught during the Fall semester 2017 at Carnegie Mellon University. My goal for this course is to give you a reasonable introduction to several fundamental areas of combinatorics. My rough plan is to split the course into four modules each lasting three to four weeks. The four main topics covered will be: enumeration, probabilistic methods, extremal graph theory, and algebraic methods in combinatorics.

I have laid out an ambitious schedule for this course, and it will be hard. You should expect to spend many hours a week reading the texts, reworking your notes, or doing homework problems. You should expect that you will spend a long time on a single problem and possibly still not be able to solve it. This is ok! My goal is for you to learn as much math as possible, and the best way to learn mathematics is to do mathematics. Even though the course will be hard, as long as you show me that you have learned a lot of mathematics, I will grade reasonably.

Course announcements will be found at

<http://www.math.cmu.edu/~mtait/21-301Fall2017>

and you should check this page regularly.

# 2

## Enumeration

*There is no problem in all mathematics that cannot be solved by direct counting.*

– Ernst Mach

### 2.1 Sequences and the Multiplication Principle

As you may have realized in calculus, the study of sequences is quite important in mathematics, and this is where we will start our course. In this lecture, we will use the terms *sequence*, *string*, and *word* interchangeably. From calculus, you know that a sequence is just an ordered list of objects, and we will now make this definition precise.

We will use the notation  $[n]$  as shorthand for the set of natural numbers from 1 to  $n$ ,  $\{1, 2, \dots, n\}$ . A sequence (or string, or word) is defined to be a function  $s : [n] \rightarrow X$  where  $X$  is some fixed set of objects. We will refer to the set  $X$  as the *alphabet* or the *set of characters*. That is, for each natural number we are associating to it a unique element of  $X$ . In calculus,  $X$  was generally the set of real numbers,

and a sequence was just an ordered list of real numbers (that is, you are associating a real number to 1, a real number to 2, a real number to 3, etc).

We will sometimes write  $s = s_1s_2s_3 \dots s_n$  rather than  $s(1) = s_1, s(2) = s_2, \dots, s(n) = s_n$ . You should think of a string as an element of the  $n$ -fold cartesian product  $X \times X \times \dots \times X$ .

We will be interested in fixing some property of our string that we are interested in, and then counting how many strings satisfy that property. If  $X$  is the set of lower case letters  $\{a, b, c, \dots, y, z\}$ , then *combinatorics*, *abcdefghijklm* and *zzzzzzzzzzzzzz* are all examples of 13 letter words coming from the alphabet  $X$ . How many 13 letter words from  $X$  are there?

**Multiplication Principle.** The number of sequences  $(s_1, s_2, \dots, s_n)$  such that there are  $a_i$  choices for  $s_i$  after having chosen  $s_1, s_2, \dots, s_{i-1}$  for each  $i = 1, 2, \dots, n$  is exactly  $a_1a_2 \dots a_n$ . In particular, if  $S_1, S_2, \dots, S_n$  are finite sets, then

$$|S_1 \times S_2 \times \dots \times S_n| = \prod_{i=1}^n |S_i|.$$

By the Multiplication Principle, the number of 13 letter words coming from the alphabet of lower case letters is exactly  $26^{13}$ . We note that the Multiplication Principle does not require that each entry comes from the same base set. In our original definition of a word, we had each entry coming from some alphabet which we called  $X$ . In practice, the interesting questions will arise when there is some restriction on what the entry in the  $i$ 'th position is allowed to be. For each  $i$ , we may be constrained to pick an element from a subset  $X_i \subset X$ , and in this case each word will be an element of

$$X_1 \times X_2 \times \dots \times X_n.$$

The Multiplication Principle allows us to count the number of words we are interested in in this case as well.



**Example.** How many strings of digits from 1–9 of length 5 have an odd number in the odd positions and an even number in the even positions?

We may split the digits 1–9 into odd and even, say  $X_o = \{1, 3, 5, 7, 9\}$  and  $X_e = \{2, 4, 6, 8\}$ . Then the strings  $s_1s_2s_3s_4s_5$  which satisfy the required property must have  $s_1, s_3, s_5 \in X_o$  and  $s_2, s_4 \in X_e$ . By the Multiplication Principle, we have that the number of such strings is  $|X_o|^3|X_e|^2 = 5^3 \cdot 4^2$ .

Sometimes it is tricky to count the number of sequences we are interested in all at once, but more straightforward once break the problem into smaller, disjoint pieces.

**Example.** Rob has 4 blue socks, 7 red socks, 5 white socks, and 3 black socks. Rob likes to wear either a red sock on his left foot with a blue sock on his right foot or a white sock on his left foot with a black sock on his right foot. How many ways are there for Rob to choose his socks?

By the Multiplication Principle, there are  $4 \cdot 7 = 28$  ways for Rob to wear a blue and red sock, and there are  $5 \cdot 3 = 15$  ways for him to wear a white and a black sock. How many are there total? It is clear that we should add 28 and 15.

**Summation Principle.** *If  $S_1, S_2, \dots, S_n$  are finite disjoint sets, then*

$$\left| \bigcup_{i=1}^n S_i \right| = |S_1| + |S_2| + \dots + |S_n|.$$

## 2.2 Permutations and Combinations

What if we want a sequence of integers from  $[n]$  of length  $k$  where all of the entries are distinct? There is no restriction for the first element of the string, ie there are  $n$  choices. When choosing the second element of the string we can choose any integer in  $[n]$  except for the first number in the sequence, so there are  $n - 1$  choices. Continuing

this process we see that there are

$$n(n-1)(n-2)\cdots(n-k+1)$$

sequences in which all elements are distinct. We call such a sequence a *permutation*.

For a positive integer  $n$ , we define  $n$  factorial, written  $n!$ , as

$$n! := n \cdot (n-1) \cdots 2 \cdot 1.$$

We denote the number of permutations of length  $k$  coming from alphabet  $[n]$  as  $P(n, k)$ . By the above discussion

$$P(n, k) = n(n-1)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

**Example.** A website requires that your password be a string of 7 lower case letters with no repeated letters. Then there are  $P(26, 7) = 26 \cdot 25 \cdots 20$  possible passwords.

Note that when trying to figure out how many distinct passwords there are, the order of the characters matters. In general in a permutation the order of the letters of your string matters.

**Example.** A class of 100 students is electing a president, a vice president, and a treasurer. Then there are  $P(100, 3)$  ways that the leadership positions can be filled.

Again, here the order of the choices matters (ie if Bob is president and Alice is vice president, then it is a different leadership than if Alice is president and Bob is vice president). Let's look at two similar examples.

**Example.** You are playing Scrabble and there is exactly one of each letter left in the bag. You are about to draw 7 tiles. How many different hands can you draw?

**Example.** A class of 100 students is electing 3 students to fill a leadership committee. How many different ways can they choose their leadership?

In the Scrabble example, if I choose  $\{A, B, C, D, E, F, G\}$  it is the same hand as if I choose  $\{G, A, B, F, C, E, D\}$ . In the leadership example, electing Alice, Bob, and Eve to the committee is the same as electing Eve, Alice, and Bob to the committee. That is, in contrast to the permutation examples, in these two cases the order of choices does not matter.

Let  $S$  be a finite set with  $|S| = n$ , and let  $k$  be a positive integer with  $k \leq n$ . We will use the notation  $\binom{n}{k}$  (said:  $n$  choose  $k$ ) to denote the number of ways to choose  $k$  (unordered) elements from  $S$ .

In the Scrabble example, we could take each unordered hand and turn it into many different 7 letter passwords of distinct letters. How many? By the Multiplication Principle we could order each set of 7 unordered letters into  $7!$  ordered sequences. Similarly, there are  $3!$  choices of President, Vice President, and Secretary for each unordered 3-person committee.

Using these ideas we can see that  $\binom{n}{k}k! = P(n, k)$ . To see this, note that if we were trying to count  $P(n, k)$  we could first choose the  $k$  elements to be in our permutation, then we could order them. This gives us the formula

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

## 2.3 Bijections and Double Counting

*My profit's so nice me and dog gon' count it twice.*

– Rick Ross

We may look at the formula for  $\binom{n}{k}$  and see that

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

Can we explain why this is true in a way other than algebraic manipulation? That is, can we give a combinatorial reason why this should be true?

The binomial coefficient  $\binom{n}{k}$  counts the number of ways to distinguish  $k$  elements in an  $n$  element set. When we are setting aside  $k$  elements which we are “choosing”, we are also setting to the other side  $n - k$  elements which we are “not choosing”. Therefore, choosing  $k$  elements is equivalent to not choosing  $n - k$  elements. Formally, what we have done is created a bijection between the set of all  $k$ -element subsets and the set of all  $n - k$  element subsets, the function being defined by taking the complement. If there is a one-to-one correspondence between two finite sets, then they must have the same size. This is the Principle of Bijection.

**Bijection Principle.** If  $S$  and  $T$  are two finite sets and there is a bijection from  $S$  to  $T$ , then  $|S| = |T|$ .

Let’s see a few examples of using the Bijection Principle. The first one is so natural that it is almost trivial.

**Example.** There is a natural bijection between binary strings of length  $n$  (ie sequences of length  $n$  coming from the alphabet  $\{0, 1\}$ ) and subsets of an  $n$  elements set. Specifically, let  $X = \{x_1, \dots, x_n\}$  be a set of  $n$  elements. Then given a binary string  $s = s_1s_2 \dots s_n$ , create a subset of  $X' \subset X$  by putting  $x_i$  in the subset if and only if  $s_i = 1$ .  $s$  is sometimes called the *characteristic vector* for  $X'$ .

This example shows that the number of binary strings of length  $n$  with exactly  $k$  1s is the same as the number of ways to choose  $k$  elements from an  $n$ -element set,  $\binom{n}{k}$ . Let’s combine this with our first use of *double counting*.

**Double Counting Principle.** If two expressions count the same set, then they must be equal.

**Example.** Prove the following identity

$$\sum_{k=0}^n \binom{n}{k} = 2^n.$$

Let's count the number of binary strings of length  $n$  in two different ways. Going through each position of a binary string, there is a choice of whether that position should be a 0 or a 1. Since there are  $n$  positions, by the Multiplication Principle there are  $2^n$  binary strings of length  $n$ . Now let's count binary strings based on the number of 1s there are. By the previous example there are  $\binom{n}{k}$  binary strings with  $k$  1s. A binary string can have between 0 and  $n$  1s inclusive. By the Summation Principle there are  $\sum_{k=0}^n \binom{n}{k}$  binary strings of length  $n$ .

**Homework.** Prove the identity

$$\sum_{k=0}^n \binom{n}{k} 6^k = 7^n.$$

**Example.**

$$\sum_{k=0}^n \binom{n}{k}^2 = \binom{2n}{n}.$$

*Proof.*  $\binom{2n}{n}$  is the number of binary strings of length  $2n$  with exactly  $n$  1s. Let's count these strings in another way. Note that if a binary string of length  $2n$  has exactly  $n$  1s, then if it has  $k$  1s in the first  $n$  positions it must have exactly  $n - k$  1s in the second  $n$  positions. Thus we can enumerate binary strings of length  $2n$  with exactly  $n$  1s by partitioning them according to how many 1s are in the first  $n$  positions. By the Summation Principle, the number of them is exactly

$$\sum_{k=0}^n \binom{n}{k} \binom{n}{n-k}.$$

By the identity  $\binom{n}{k} = \binom{n}{n-k}$  we are done. □

**Theorem** (Pascal's Identity). Define  $\binom{n}{0}$  to be 1. Then for  $n, k \geq 0$

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

*Proof.* We can count  $k$ -subsets of an  $n$  element set in two different ways. Enumerating them directly gives  $\binom{n}{k}$ . Let the “first” element of the  $n$ -element set be distinguished. Then we can partition the  $k$  element subsets into subsets that have the first element and those that don't. There are  $\binom{n-1}{k}$   $k$ -subsets which do not contain the first element and there are  $\binom{n-1}{k-1}$   $k$ -subsets which do contain the first element.  $\square$

Next we will consider another natural bijection between a combinatorial object and subsets. Consider an  $(x \times y)$  integer lattice of points in  $\mathbb{Z}^2$ . Assume that the bottom left hand corner of the rectangle is the origin and the top right hand corner is the point  $(x, y)$ . An  $(x, y)$  *lattice path* is a sequence of unit steps up and to the right, starting at the origin and ending at  $(x, y)$ . We let  $L(x, y)$  be the number of  $(x, y)$  lattice paths. It is clear that  $L(x, 0) = L(0, y) = 1$  for any natural numbers  $x$  and  $y$ , and we will define  $L(0, 0)$  to be 1.

There is a natural bijection between  $(x, y)$  lattice paths and binary strings of length  $x + y$ . Namely, given an  $(x, y)$  lattice path, we construct a binary string of length  $x + y$  where the  $i$ 'th position is  $R$  if the  $i$ 'th step of the path is to the right and  $U$  if the  $i$ 'th step of the path is up. By the Bijection Principle, this gives

$$L(x, y) = \binom{x+y}{x} = \binom{x+y}{y}.$$

**Example.** Classifying lattice paths by whether their first step is up or to the right gives the recurrence  $L(x, y) = L(x-1, y) + L(x, y-1)$ . This is Pascal's identity.

**Homework.** How many  $(n, n)$  lattice paths are there that never go below the line  $y = x$ ?

**Homework.** How many subsets of  $[n]$  are there with odd size?

We will now discuss *compositions*, which are counted by binomial coefficients despite not appearing to be counted by sets. A composition of  $n$  is a sequence of positive integers  $(x_1, \dots, x_k)$  which sum to  $n$ . Each  $x_i$  is called a *part*. How many compositions of  $n$  are there into  $k$  parts? One way to think of this is instead of choosing each part, you can choose what the partial sums are. Since the  $k$  parts must sum to  $n$ , once  $x_1, \dots, x_{k-1}$  is chosen  $x_k$  is determined. Therefore the composition is determined by the first  $k-1$  partial sums of the parts. Any  $k-1$  distinct numbers in  $[n-1]$  are possible partial sums, and any  $k-1$  partial sums can be reverse engineered into a unique list of  $k-1$  parts. Therefore the number of compositions of  $n$  into  $k$  parts is exactly  $\binom{n-1}{k-1}$ .

**Example.** A mother has 23 one dollar bills that she wants to distribute between her 4 children such that each child receives at least 1 dollar. In how many ways can she do this?

We note that in this example we are assuming that the children are distinct (their order matters), but the one dollar bills are indistinguishable (their order does not matter). We can reformulate this problem into an integer programming problem. We want the number of integer solutions to

$$x_1 + x_2 + x_3 + x_4 = 23$$

subject to the constraint  $x_i \geq 1$ . This is exactly a composition of 23 into 4 parts! So the number of ways is  $\binom{22}{3}$ . What if we drop the assumption that each child receives at least one dollar? Then we have changed the constraints to  $x_i \geq 0$ . To solve this problem we may create new variables  $z_i = x_i + 1$ . Then the equation becomes

$$(z_1 - 1) + (z_2 - 1) + (z_3 - 1) + (z_4 - 1) = 23$$

subject to constraints  $z_i \geq 1$ . The rest is left as an exercise.

We end this section with the Binomial Theorem.

**Theorem** (Binomial Theorem). *Let  $n$  be a natural number.*

$$(x + y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}.$$

*Proof.* Expand the left hand side, and classify according to the number of  $x$ 's taken from the factors. □

Note that the binomial theorem immediately gives the identity  $\sum_{k=0}^n \binom{n}{k} = 2^n$  by setting  $x = y = 1$ . What other identities can you prove with it?

## 2.4 Estimation

*log log log  $n$  is proven to go to infinity with  $n$ , but has never been observed to do so.*

– Carl Pomerance

In the previous sections, we were able to count certain quantities exactly. This is the ideal scenario, but is not always possible. Sometimes it is too difficult to count something exactly (eg, for very large  $n$ , how many prime numbers are there between 1 and  $n$ ?). When this is the case, what we as mathematicians should try to do is obtain the best upper and lower bounds we can for the quantity in question.

Being able to estimate quantities quickly and effectively is an important skill in combinatorics. With time you will get a feel for when terms can be thrown away because they are not asymptotically important, for when care needs to be taken with error terms, for when it is appropriate to obtain a crude bound on a quantity rather than a precise one, etc. Often we estimate quantities by simple and often-used functions like log, polynomials, exponentials. When doing so, it is important to know “how fast” functions grow. For example, you may remember from calculus (say, using



L'Hopital's rule) that for any  $\epsilon > 0$ ,  $n^\epsilon$  is “much bigger” than  $\log n$  and for any  $\alpha > 1$ ,  $x^\alpha$  is “much bigger” than any polynomial in  $n$ . This can be made precise using limits. In mathematics and computer science, we use “Big-Oh” notation to succinctly convey how large one function is compared to another.

Notation	Definition	Meaning
$f(n) = O(g(n))$	$\limsup_{n \rightarrow \infty} \frac{f(n)}{g(n)}$ is finite	$f$ not much bigger than $g$
$f(n) = \Omega(g(n))$	$g(n) = O(f(n))$	$f$ not much smaller than $g$
$f(n) = \Theta(g(n))$	$f(n) = O(g(n))$ $f(n) = \Omega(g(n))$	$f$ and $g$ have the same order of magnitude
$f(n) = o(g(n))$	$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 0$	$f$ is much smaller than $g$
$f(n) \sim g(n)$	$\lim_{n \rightarrow \infty} \frac{f(n)}{g(n)} = 1$	$f$ and $g$ grow at the same rate

Let's try to estimate a nontrivial function that appears often in mathematics and computer science. Let

$$H_n := 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} = \sum_{k=1}^n \frac{1}{k}.$$

$H_n$  is called the  $n$ th *harmonic number*. You learned in calculus that  $H_n \rightarrow \infty$  as  $n \rightarrow \infty$ , but how fast does it go to infinity? One way is to group the terms as

$$H_n = 1 + \left(\frac{1}{2} + \frac{1}{3}\right) + \left(\frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7}\right) + \cdots.$$

That is, we are grouping together the numbers  $\frac{1}{k}$  where

$$\frac{1}{2^k} < \frac{1}{k} \leq \frac{1}{2^{k-1}}.$$

The number of terms in each subsequent set of parentheses will double and the number of total pairs of parentheses will be  $1 + \lfloor \log_2 n \rfloor$ . Next we notice that in the  $i$ 'th set of parentheses, there are  $2^i$  terms each of which are at most  $\frac{1}{2^i}$ . This means that  $H_n$

is bounded above by the number of pairs of parentheses. Similarly, besides perhaps the last set of parentheses, each set contains  $2^i$  terms each of which are strictly greater than  $\frac{1}{2^{i+1}}$ . Therefore  $H_n$  is bounded below by  $\frac{1}{2}$  times the number of pairs of parentheses minus 1. We have shown

$$\frac{1}{2} \lfloor \log_2 n \rfloor < H_n \leq 1 + \log_2 n.$$

This shows that  $H_n = \Theta(\log n)$ , so we have determined its order of magnitude.

**Homework.** Use calculus to show that  $H_n \sim \log_2 n$ .

Let's now try to estimate the factorial function. Some first attempts are as follows.

$$2^{n-1} = \prod_{i=2}^n 2 \leq \prod_{i=1}^n i = n! = \prod_{i=1}^n i \leq \prod_{i=1}^n n = n^n.$$

So  $n!$  is somewhere between  $2^{n-1}$  and  $n^n$ . Is either of these estimates any good? Being a bit more careful (and ignoring ceilings and floors), we can see that

$$n! = \prod_{i=1}^{n/2} i \prod_{i=n/2+1}^n i \leq \left(\frac{n}{2}\right)^{n/2} n^{n/2} = \frac{n^n}{2^{n/2}}.$$

So  $n! = o(n^n)$ . On the other hand, for all  $1 \leq i \leq \frac{n}{2}$  we have  $i(n-i+1) \geq n$  (why?), which gives that (assume  $n$  is even for simplicity)

$$n! = \prod_{i=1}^{n/2} i(n-i+1) \geq n^{n/2}$$

Now  $2^{n-1} = o(n^{n/2})$  (why?), so neither of our estimates is good so far. We can be much more precise:

**Theorem.** For  $n \geq 1$  we have

$$e \left(\frac{n}{e}\right)^n \leq n! \leq en \left(\frac{n}{e}\right)^n.$$

*Proof.* You will prove the lower bound in your first homework. For the upper bound, we will consider  $\ln n!$ . By properties of the logarithm, we have

$$\ln n! = \ln 1 + \ln 2 + \cdots + \ln n$$

We note that this is a right Riemann sum for the function  $f(x) = \ln x$  on the interval from 1 to  $n + 1$ . Since  $f(x)$  is concave down, the right Riemann sum underestimates the area under the curve. Therefore,

$$\ln n! \leq \int_1^{n+1} \ln x \, dx = (n + 1) \ln(n + 1) - n.$$

Now

$$n! = e^{\ln n!} \leq e^{(n+1)\ln(n+1)-n} = \frac{(n+1)^{n+1}}{e^n}.$$

This is equivalent to the expression in the upper bound. □

We use this to give simple but very useful upper and lower bounds for the binomial coefficient.

**Theorem.** For  $n \geq 1$  and  $1 \leq k \leq n$ , we have

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k.$$

*Proof.* Fix  $n$  and  $k$  with  $k \leq n$ . Calculus shows that the function  $f(x) = \frac{n-x}{k-x}$  is nondecreasing. Therefore, for any  $0 \leq i \leq k - 1$ , we have  $\frac{n}{k} \leq \frac{n-i}{k-i}$ . Then

$$\binom{n}{k} = \frac{n}{k} \frac{n-1}{k-1} \cdots \frac{n-k+1}{k-k+1} \geq \left(\frac{n}{k}\right)^k.$$

On the other hand

$$\binom{n}{k} \leq \frac{n^k}{k!}.$$

Using the previous theorem gives the upper bound. □

## 2.5 Inclusion-Exclusion

In our section of Math 301, there are 32 students, 21 math majors, and 14 seniors. How many students are there that are neither math majors nor seniors? This is equivalent to counting the number of students who are either math majors or seniors (or both). It is clear that we do not have enough information to solve this problem. We can't simply add the number of math majors to the number of seniors, because we will over count those students that are senior math majors. However, it is clear that if we add 21 to 14, then we have counted people who are either math majors or seniors but not both exactly once, and we have counted senior math majors exactly twice. Therefore, if we subtract the number of senior math majors from  $21 + 14$  we will have found the number of students who fit into at least one of the two categories. Our class has 3 students who are senior math majors, which means that the number of students who are neither math majors nor seniors is

$$32 - (21 + 14 - 3) = 2.$$

This example shows us how to count when there are two sets: if  $A$  and  $B$  are finite sets, then

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

What if there are three sets? Let  $A$ ,  $B$ , and  $C$  be finite sets. Then

$$|A| + |B| + |C|$$

counts every element in  $A \cup B \cup C$ , but it counts elements that are in more than one set more than once. Similar to before we may try

$$|A| + |B| + |C| - |A \cap B| - |B \cap C| - |A \cap C|.$$

Now any element in exactly one of the sets is counted exactly once, and any element in exactly two of the sets is counted exactly  $2 - 1$  times. But if an element is in all three sets it is counted  $3 - 3$  times. Therefore, we see that

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|.$$

Now try 4 sets and then 5. You may see a pattern emerging. It will be more convenient for us to count the number of elements which are in *none* of the sets, which is the same as taking the number of elements in the union and taking the complement.

**Theorem** (Inclusion-Exclusion Theorem). *Let  $A_1, \dots, A_m$  be a collection of  $m$  subsets of a finite set  $X$ . Then*

$$\left| X \setminus \left( \bigcup_{i=1}^m A_i \right) \right| = \sum_{S \subseteq [m]} (-1)^{|S|} \left| \bigcap_{i \in S} A_i \right|$$

where we define the intersection of zero sets to be  $X$ .

*Proof.* If an element of  $X$  is in none of the sets  $A_1, \dots, A_m$ , then it is counted exactly once on the left hand side, and in the sum on the right hand side it is included exactly once, when  $S = \emptyset$ . Assume that an element of  $X$  is included in exactly  $n$  of the sets in  $A_1, \dots, A_m$  where  $n > 0$ . Then the number of subsets  $S$  with  $|S| = k$  such that the element is included in  $\bigcap_{i \in S} A_i$  is exactly  $\binom{n}{k}$ . Therefore, the number of times the element is counted in the sum on the right hand side is exactly

$$\sum_{k=0}^n (-1)^k \binom{n}{k}.$$

By the binomial theorem, this is equal to 0. □

Let's see some examples. The first three are classical.

**Example** (Counting Derangements). A sequence of length  $n$  from  $[n]$  which is a bijection is called a *permutation*. We use the notation  $S_n$  to denote the set of all permutations of  $[n]$ . A permutation  $\sigma \in S_n$  is called a *derangement* if  $\sigma(i) \neq i$  for all  $i = 1, 2, \dots, n$ . How many derangements of  $[n]$  are there?

We can solve this problem by using the Inclusion-Exclusion Theorem. We will let the set  $A_i \subset S_n$  be the set of permutations  $\tau$  such that  $\tau(i) = i$ . Then a permutation  $\sigma$  is a derangement if it is not in any of the sets  $A_i$ . Let  $S \subset [n]$  have size  $k$ . To apply the theorem we must be able to calculate the size of

$$\bigcap_{i \in S} A_i.$$

A permutation is in this intersection if and only if it is fixed on the set of  $k$  indices of  $S$ . The permutation may do anything else outside of those indices (on  $n - k$  other positions). Therefore,

$$\left| \bigcap_{i \in S} A_i \right| = (n - k)!.$$

The number of subsets of  $[n]$  of size  $k$  is exactly  $\binom{n}{k}$ , therefore the Inclusion-Exclusion Theorem gives us that

$$\left| S_n \setminus \left( \bigcup_{i=1}^n A_i \right) \right| = \sum_{k=0}^n (-1)^k \binom{n}{k} (n - k)!.$$

This is a good result, but we should try to simplify the formula. Let  $\mathcal{D}$  be the set of all derangements.

$$|\mathcal{D}| = \sum_{k=0}^n (-1)^k \binom{n}{k} (n - k)! = \sum_{k=0}^n (-1)^k \frac{n!}{k!} = n! \sum_{k=0}^n (-1)^k \frac{1}{k!}.$$

Recalling from calculus that the Taylor series for  $e^x$  is

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!},$$

We see that the sum is a truncated Taylor series for  $\frac{1}{e}$ . This tells us that as  $n \rightarrow \infty$ , the number of derangements of  $[n]$  tends to  $\frac{1}{e}$  of the total number of permutations of  $[n]$ .

**Example** (Counting Surjections). Let  $m \leq n$ . How many surjections are there  $f : [n] \rightarrow [m]$ ?

A function  $f$  is not a surjection if there is a  $k \in [m]$  such that  $f(i) \neq k$  for all  $i \in [n]$ . Therefore, let  $A_k$  be the set of functions  $f : [n] \rightarrow [m]$  such that  $k$  is not in the range of  $f$ . Let  $\mathcal{F}$  be the set of all functions from  $[n] \rightarrow [m]$ . Then

$$\left| \mathcal{F} \setminus \left( \bigcup_{k=1}^m A_k \right) \right|$$

is exactly the number of surjections from  $[n]$  to  $[m]$  (let's call this set  $S(n, m)$ ). Then Inclusion-Exclusion tells us

$$|S(n, m)| = \sum_{i=0}^m (-1)^i \binom{m}{i} \left| \bigcap_{\substack{k \in S \\ |S|=i}} A_k \right|.$$

We must count the size of

$$\bigcap_{\substack{k \in S \\ |S|=i}} A_k.$$

Functions in this intersection have a fixed set of  $i$  values in  $[m]$  which are not in the range of the function. Therefore, any function from  $[n]$  to  $[m] \setminus S$  is in the intersection. There are  $(m - i)^n$  of these functions. So

$$|S(n, m)| = \sum_{i=0}^m (-1)^i \binom{m}{i} (m - i)^n.$$

**Example** (The Euler  $\phi$  function). Let  $n \geq 2$  be a positive integer. The *Euler totient function* is defined by

$$\phi(n) = |\{m \in \mathbb{N} : m \leq n, \gcd(m, n) = 1\}|.$$

If  $n$  has prime factorization  $n = p_1^{e_1} \cdots p_m^{e_m}$ , what is  $\phi(n)$ .

We want to count the number of natural numbers less than  $n$  which are coprime to  $n$ . If  $n = p_1^{e_1} \cdots p_m^{e_m}$  then a number is coprime with  $n$  if and only if it is not divisible by  $p_i$  for any  $i \in [m]$ . Therefore, if we let  $A_i$  be the set of natural numbers in  $[n]$  which are divisible by  $p_i$ , then

$$\phi(n) = n - \left| \bigcup_{i=1}^m A_i \right|.$$

To count

$$\left| \bigcap_{i \in S} A_i \right|,$$

let  $S = \{p_{i_1}, \dots, p_{i_k}\}$ . Then a number  $r \in \bigcap_{i \in S} A_i$  if and only if  $p_{i_1} \cdots p_{i_k} | r$ . The number of  $r \leq n$  which satisfy this property is exactly

$$\frac{n}{p_{i_1} \cdots p_{i_k}}$$

(viz, if  $P = p_{i_1} \cdots p_{i_k}$  then they are exactly the integers  $\{P, 2P, \dots, (n/(p_{i_1} \cdots p_{i_k}))P\}$ ).

Therefore, by Inclusion-Exclusion

$$\phi(n) = \sum_{S \subset [m]} (-1)^{|S|} \frac{n}{\prod_{i \in S} p_i} = n \sum_{S \subset [m]} \frac{(-1)^{|S|}}{\prod_{i \in S} p_i} = n \prod_{i=1}^m \frac{p_i - 1}{p_i}.$$

(Why is the last equality true?).

## 2.6 Generating Functions

*This is the second magic of generating functions: the generating function for complicated things can be obtained from the generating function for simple things.*

– Michel Goemans



In this section we discuss a new way to talk about sequences. At first, this is just a change of notation, but we will see that this change of notation allows us to use powerful analytic methods. Given a sequence  $\{a_n\}$ , we may construct a *generating function*

$$A(x) = \sum a_n x^n.$$

Notice that we have not gained or lost any information here, we are simply associating the  $i$ 'th element of the sequence with a symbol  $x^i$ . Wilf describes a generating function as “a clothesline on which we hang up a sequence of numbers for display”. Often times we are considering a class of combinatorial objects  $\mathcal{A}$  to be enumerated. In this case, we call  $A(x)$  the generating function for  $\mathcal{A}$

$$A(x) = \sum a_n x^n$$

where  $a_n$  is the number of objects in the class that have size (or weight)  $n$  (here we are assuming that the number of objects in  $\mathcal{A}$  of each fixed size is finite). Let's look at some simple examples.

**Example.** The sequence  $\{a_n\} = \{1, 1, 1, 1, \dots\}$  has generating function

$$A(x) = \sum_{n=0}^{\infty} x^n = 1 + x + x^2 + x^3 + \dots.$$

Note that from calculus, if  $|x| < 1$ , then  $A(x) = \frac{1}{1-x}$ . (This is a geometric series).

**Example.** Let  $\mathcal{A}$  be the set of binary sequences. For a binary sequence, let its size be the length of the sequence. Then the generating function for  $\mathcal{A}$  is

$$A(x) = \sum_{n=0}^{\infty} 2^n x^n.$$

**Example.** Let  $\mathcal{A}$  be the set of subsets of an  $N$  element set. For a subset, let its size be the number of elements in the subset. Then the generating function for  $\mathcal{A}$  is

$$A(x) = \sum_{n=0}^N \binom{N}{n} x^n.$$

Note that for now we are treating the “variable”  $x$  just as a placeholder. If we thought of it as a function, we could write the previous generating function as  $A(x) = (1 + x)^N$ . Let’s see some examples where we work with the generating function.

**Example.** The Fibonacci numbers  $F_n$  are defined recursively.  $F_1 = F_2 = 1$ , and for  $n \geq 2$ ,  $F_n = F_{n-1} + F_{n-2}$ . The generating function for this sequence is

$$A(x) = \sum_{n=0}^{\infty} F_n x^n.$$

(define  $F_0 = 0$ ). Multiplying the recurrence  $F_n = F_{n-1} + F_{n-2}$  by  $x^n$  and summing from  $n \geq 2$  gives

$$\sum_{n \geq 2} F_n x^n = \sum_{n \geq 2} F_{n-1} x^n + \sum_{n \geq 2} F_{n-2} x^n.$$

Notice that  $\sum_{n \geq 2} F_n x^n = A(x) - x$ . Also notice that  $\sum_{n \geq 2} F_{n-1} x^n = xA(x)$  and  $\sum_{n \geq 2} F_{n-2} x^n = x^2 A(x)$ . Therefore, we see

$$A(x) = \frac{x}{1 - x - x^2}.$$

Our goal is to obtain the coefficients  $F_n$ , so we may try to simplify by using the method of partial fractions. Let  $\phi = \frac{1+\sqrt{5}}{2}$  and  $\bar{\phi} = \frac{1-\sqrt{5}}{2}$ . Then

$$\begin{aligned} A(x) &= \frac{-x}{(x - \phi)(x - \bar{\phi})} = \frac{1}{\sqrt{5}} \left( \frac{1}{1 - \phi x} - \frac{1}{1 - \bar{\phi} x} \right) \\ &= \frac{1}{\sqrt{5}} \left( (1 + \phi x + \phi^2 x^2 + \dots) - (1 + \bar{\phi} x + \bar{\phi}^2 x^2 + \dots) \right). \end{aligned}$$

So we can see that the coefficient of  $x^n$  is  $\frac{1}{\sqrt{5}}(\phi^n - \bar{\phi}^n)$ , which gives an explicit formula for  $F_n$ .

Were you uncomfortable with that calculation? The recurrence is a definition, but why are we allowed to take an infinite sum on both sides? Then we did algebraic

manipulations (viz the method of partial fractions and the expansion of a geometric series) on something where I told you that the variable was just a placeholder. Does this make any sense? Let's formalize this.

## 2.7 Formal Power Series

Let  $\{a_n\}_{n \geq 0}$  be any sequence of real numbers. For this sequence, we formally define an object (which we call a power series)

$$A(x) = \sum_{n=0}^{\infty} a_n x^n.$$

Now we let  $R[x]$  be the set of all such  $A(x)$  as  $\{a_n\}$  runs over all sequences of real numbers. As of now, this is just a set of objects. We call  $R[x]$  the ring of formal power series. We will now define two operations on  $R[x]$  as follows. If  $A(x) = \sum a_n x^n$  and  $B(x) = \sum b_n x^n$ , then

$$A(x) + B(x) = \sum_{n=0}^{\infty} (a_n + b_n) x^n$$

and

$$A(x) \cdot B(x) = \sum_{n=0}^{\infty} c_n x^n,$$

where  $c_n = \sum_{k=0}^n a_k b_{n-k}$ . One can check (using properties of real numbers), that  $R[x]$  satisfies all of the axioms of a ring, where the multiplicative identity is the function  $A(x) = 1 + \sum_{n \geq 1} 0x^n$  and the additive identity is  $\sum_{n \geq 0} 0x^n$ . One can also check that this is an integral domain. We define the *coefficient operator* for each  $n$  to be

$$[x^n]A(x) = a_n.$$

That is, it extracts the coefficient of  $x^n$  (this will be useful when asking questions like, how many objects of weight  $n$  are there in  $\mathcal{A}$ ?). Since  $R[x]$  is a ring, we may ask which elements have inverses. You should prove the following theorem as homework.

**Theorem.** Let  $A(x) = \sum_{n \geq 0} a_n x^n$ . Then  $A(x)$  has an inverse if and only if  $a_0 \neq 0$ .

In the case that  $a_0 \neq 0$ , then there is a unique  $B(x) \in R[x]$  such that  $A(x) \cdot B(x) = 1$ . In this case we write

$$A(x) = \frac{1}{B(x)}.$$

One may now check formally that the function  $A(x) = 1 - x$  and the function  $B(x) = \sum_{n \geq 0} x^n$  are inverses in  $R[x]$ . Our intuition from geometric series carries over to this ring. Maybe other properties that we “think should be true” will be. We can use the geometric series to compute partial sums: if  $\{a_n\}$  is a sequence with generating function  $A(x)$ , and  $b_n = \sum_{k=0}^n a_k$  is the sequence of partial sums, then the generating function for  $\{b_n\}$  is given by

$$B(x) = \frac{A(x)}{1 - x}.$$

This is checked formally by the definition of multiplication in  $R[x]$ . We also define a formal derivative and integral on  $R[x]$ , which take a power series as input and output another power series.

$$\frac{\partial}{\partial x} \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} n a_n x^{n-1}$$

and

$$\int \sum_{n=0}^{\infty} a_n x^n = \sum_{n=0}^{\infty} \frac{a_n}{n+1} x^{n+1}.$$

For a real number  $r$  we define the formal power series

$$(1+x)^r = \sum_{k=0}^{\infty} \binom{r}{k} x^k.$$

This generalizes the binomial theorem when  $r$  is a positive integer. One can check

using the formal definition of addition and multiplication that in  $R[x]$  the usual rules

$$\begin{aligned}(1+x)^a(1+x)^b &= (1+x)^{a+b} \\ ((1+x)^a)^b &= (1+x)^{ab} \\ (1+x)^{-a} &= \frac{1}{(1+x)^a}\end{aligned}$$

hold. One can also check that the formal power series  $(1+x)^r$  matches the Taylor series for  $(1+x)^r$  from calculus.

## 2.8 Generating Functions Redux

We will most often be interested in using generating functions to count combinatorial objects. For example, we previously saw that the generating function for binary strings is given by

$$A(x) = \sum_{n \geq 0} 2^n x^n.$$

This generating function gives us the information that the number of binary strings of length  $n$  is  $2^n$ . Note here two important details: the length of a binary string is well-defined and we are grouping together strings with the same length, and the number of strings of a fixed length is finite. In general, we will be interested in counting objects with these two properties.

**Definition.** A *set of combinatorial objects* is a finite or countably infinite set on which a size function is defined satisfying

1. The size of an element is a nonnegative integer.
2. For any  $k \in \mathbb{N} \cup \{0\}$ , the number of elements in the set of size  $k$  is finite.

If  $\mathcal{A}$  is a set of combinatorial objects, we may define the sequence  $\{a_n\}$  where  $a_n$  is the number of elements of  $\mathcal{A}$  with size  $n$ . This gives us a natural generating function

$$\Phi_{\mathcal{A}}(x) = \sum_{n \geq 0} a_n x^n = \sum_{\alpha \in \mathcal{A}} x^{|\alpha|},$$

where the notation  $|\alpha|$  denotes the size of  $\alpha$ . We will use the notation  $\Phi_{\mathcal{A}}(x)$  whenever we are working with the generating function of a set of combinatorial objects  $\mathcal{A}$ .

The goal of formalizing sequences that count combinatorial objects in this way is that now we can rigorously state decomposition theorems which allow us to count the number of objects by breaking them down into smaller pieces.

**Theorem** (Sum Lemma). *Let  $\mathcal{A}$  and  $\mathcal{B}$  be sets of combinatorial objects and let  $|\cdot|$  be a size function defined on  $\mathcal{A} \cup \mathcal{B}$ , then*

$$\Phi_{\mathcal{A} \cup \mathcal{B}} = \Phi_{\mathcal{A}} + \Phi_{\mathcal{B}} - \Phi_{\mathcal{A} \cap \mathcal{B}}$$

*Proof.* By the inclusion-exclusion theorem,

$$\sum_{\alpha \in \mathcal{A} \cup \mathcal{B}} x^{|\alpha|} = \sum_{\alpha \in \mathcal{A}} x^{|\alpha|} + \sum_{\alpha \in \mathcal{B}} x^{|\alpha|} - \sum_{\alpha \in \mathcal{A} \cap \mathcal{B}} x^{|\alpha|}.$$

□

**Theorem** (Product Lemma). *Let  $\mathcal{A}$  and  $\mathcal{B}$  be sets of combinatorial objects with weight functions  $|\cdot|_{\mathcal{A}}$  and  $|\cdot|_{\mathcal{B}}$ . Define a weight function on  $\mathcal{A} \times \mathcal{B}$  as  $|(a, b)| = |a|_{\mathcal{A}} + |b|_{\mathcal{B}}$ . Then*

$$\Phi_{\mathcal{A} \times \mathcal{B}} = \Phi_{\mathcal{A}} \cdot \Phi_{\mathcal{B}}.$$

*Proof.* First, it is clear that  $\mathcal{A} \times \mathcal{B}$  is a set of combinatorial objects under weight function  $|\cdot|$ , so  $\Phi_{\mathcal{A} \times \mathcal{B}}$  is well-defined. Now, if  $(a, b) \in \mathcal{A} \times \mathcal{B}$  such that  $|(a, b)| = n$ , then there must be a  $k \in \{0, \dots, n\}$  such that  $|a|_{\mathcal{A}} = k$  and  $|b|_{\mathcal{B}} = n - k$ . Counting the number of elements for which this occurs and applying the definition of multiplication of power series gives the result. □

We may use the Sum and Product Lemma in surprisingly powerful ways. We will now look at several examples where without this generating function magic it would be completely unclear how to go about counting, but with this machinery it is almost trivial.

### 2.8.1 Change making

How many ways are there to make change for a dollar using pennies, nickels, dimes, and quarters? You can probably figure this out by brute force, but let's use generating functions to do it easily and in a way that is robust to changing the constraints of the problem. Generally, we can use generating functions to find the way to make change for  $n$  cents. The key idea is to use the Product Lemma to break the problem down into pieces that are trivial to solve.

How many ways are there to make change for  $n$  cents using only pennies. Well, for each  $n$  there is exactly one way. What about using only nickels? If  $n$  is a multiple of 5 there is exactly one way, otherwise there are 0 ways. Similar reasoning is done for dimes and quarters. This gives that the generating functions for making change using pennies, nickels, dimes, and quarters are given by

$$\begin{aligned}\Phi_p(x) &= 1 + x + x^2 + \cdots = \frac{1}{1-x} \\ \Phi_n(x) &= 1 + x^5 + x^{10} + \cdots = \frac{1}{1-x^5} \\ \Phi_d(x) &= 1 + x^{10} + x^{20} + \cdots = \frac{1}{1-x^{10}} \\ \Phi_q(x) &= 1 + x^{25} + x^{50} + \cdots = \frac{1}{1-x^{25}}\end{aligned}$$

Using the product lemma, the generating function for making change with all of these coins is just the product of these functions. You can use the method of partial fractions to get an explicit formula for the number of ways to make change on  $n$  cents.

It is clear that we can also use the same method when there are other restrictions. Say for example, you are only allowed to give at most 4 pennies when you make change. Then the generating function for counting the number of ways to make change with this restriction will be

$$(1 + x + x^2 + x^3 + x^4)\Phi_n(x)\Phi_d(x)\Phi_q(x).$$

## 2.8.2 Compositions

Let's first count something that we already know how to count: compositions of  $[n]$  into  $k$  parts. It is easy to count the number of compositions of  $[n]$  into 1 part: there is exactly 1. Therefore the generating function for compositions into 1 part (where the size function is given by the sum of the parts) is just

$$\Phi = x + x^2 + x^3 + \dots = \frac{x}{1-x}.$$

Using the product lemma, the generating function for compositions into  $k$  parts is exactly

$$\Phi^k = \frac{x^k}{(1-x)^k}.$$

Using the binomial theorem, we may extract the coefficient of  $x^n$  in this series, and see that it is

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

We already knew this, but let's count something that we didn't know how to before. Suppose we want the number of compositions of  $[n]$  into  $k$  parts where every part is an odd number? The generating function for compositions into 1 part where each part (the only part) is an odd number is

$$\Phi = x + x^3 + x^5 + \dots = \frac{x}{1-x^2}.$$



Therefore the generating function for compositions of  $k$  odd integers is given by

$$\Phi^k = \left( \frac{x}{1-x^2} \right)^k.$$

Using the binomial theorem shows that the coefficient of  $x^n$  here is  $\binom{(n+k-2)/2}{(n-k)/2}$  if  $n-k$  is even and 0 if  $n-k$  is odd.

### 2.8.3 Counting Subsets

We may count sets that have restrictions on them. How many subsets of  $[n]$  of size  $k$  have no two consecutive integers? If we have such a set  $\{x_1, \dots, x_k\}$ , we may encode it as the difference between adjacent elements in the set  $(x_1, x_2 - x_1, x_3 - x_2, \dots, x_k - x_{k-1}, n+1 - x_k)$ . Then this gives a bijection between sets that we are counting and compositions of  $[n+1]$  of size  $k+1$  where every part except the first and last are at least 2. Instead of trying to find the generating function directly, let's look at each constituent part and use the Product Lemma. Here our weight function for a composition will be the sum of the integers. Then the generating function for compositions of size 1 is just

$$x + x^2 + x^3 + \dots = \frac{x}{1-x},$$

and the generating function for compositions of size 1 where each element (the only element) is at least 2 is

$$x^2 + x^3 + \dots = \frac{x^2}{1-x}.$$

Therefore the generating function for compositions with  $k+1$  parts where every party except possibly the first and last is at least 2 is

$$\frac{x}{1-x} \cdot \left( \frac{x^2}{1-x} \right)^{k-1} \cdot \frac{x}{1-x} = \frac{x^{2k}}{(1-x)^{k+1}}.$$

Using the binomial theorem, we see that the coefficient of  $x^{n+1}$  in this series is  $\binom{n-k+1}{k}$ .

## 2.8.4 Counting Strings

Let  $A$  and  $B$  be sets of strings over some fixed alphabet. We let  $AB$  denote the concatenation of  $A$  and  $B$ , which means the set of all strings  $ab$  such that  $a \in A$  and  $b \in B$ . We let  $A^*$  denote the set

$$A^* = \{\epsilon\} \cup A \cup AA \cup AAA \cup \dots$$

where  $\epsilon$  denotes the empty string. We say that the elements of  $AB$  are uniquely created if for every  $s \in AB$  there is a unique  $a \in A$  and  $b \in B$  such that  $s = ab$ . We say that elements of  $A^*$  are created uniquely similarly if all of  $\{\epsilon\}, A, AA, \dots$  are pairwise disjoint and each are uniquely created. For example  $\{0, 1\}^*$  is the set of all binary strings and is uniquely created. Then the Product and Sum lemma give the following:

**Theorem.** *Let  $A$  and  $B$  be finite sets of strings such that  $AB$  and  $A^*$  are uniquely created. Define the size of a string to be the length of that string. Then*

$$\Phi_{AB} = \Phi_A \cdot \Phi_B$$

and

$$\Phi_{A^*} = \frac{1}{1 - \Phi_A}.$$

The reason that we need the strings to be uniquely created is to ensure, for example, that  $AB$  is isomorphic to  $A \times B$ .

**Example.** Let's count the number of binary strings which do not contain consecutive zeroes (ie 00 free strings). If  $S$  is the set of strings that are 00-free, then

$$S = \{\epsilon, 0\}(\{1\}\{0, \epsilon\})^*$$

and the strings are created uniquely. Therefore, we have

$$\Phi_S(x) = (1 + x) \cdot \frac{1}{1 - x - x^2}.$$

Since the generating function for Fibonacci numbers is  $A(x) = \frac{x}{1-x-x^2}$ , we see that the number of 00-free strings of length  $n$  is  $F_{n+2}$ .



# 3

## The Probabilistic Method

*The probabilistic method is best described by examples.*

– Joel Spencer

In this chapter we will learn how to use the probabilistic method to solve combinatorial problems. The rough idea of the method is the following. Say we want to know whether a certain combinatorial configuration exists. For example, we may be given a graph and want to know whether it can be properly colored using 4 colors. One way to show that a configuration with the desired properties exists is to construct a probability space and show that a randomly chosen configuration in this space has the desired properties with non-zero probability.

### 3.1 Preliminaries

A *probability space* is a triple  $(\Omega, \mathcal{F}, \mathbb{P})$  that has

- $\Omega$  is a set.
- $\mathcal{F}$  is a subset of the power set of  $\Omega$  that contains  $\emptyset$  and that is closed under

complementation and countable unions and intersections (this is called a  $\sigma$ -field or  $\sigma$ -algebra).

- $\mathbb{P}$  is a function  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  such that  $\mathbb{P}(\Omega) = 1$  and for  $A, B \in \mathcal{F}$  that are disjoint,  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ .

Most of the time we will be interested in the case where  $\Omega$  will be a finite set and  $\mathcal{F}$  will just be the power set of  $\Omega$ . In this case, the entire probability space is determined by the values of  $\mathcal{P}$  on the elements of  $\Omega$ . These are called *elementary events* and the subsets in  $\mathcal{F}$  are called *events*. Given an event  $A$ , we say that  $\mathcal{P}(A)$  is the *probability of A*.

It is easy to show with this definition that probability is subadditive:

**Lemma 3.1.1.** *If  $A_1, \dots, A_n$  are events in a probability space, then*

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mathbb{P}(A_i).$$

*Proof.* By induction, it suffices to show that  $\mathbb{P}(A_1 \cup A_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$ . Let  $B_1 = A_1$  and  $B_2 = A_2 \setminus A_1$ . Note that  $B_1$  and  $B_2$  are disjoint, and so  $\mathbb{P}(B_1 \cup B_2) = \mathbb{P}(B_1) + \mathbb{P}(B_2)$ . Also note that  $A_2$  is the disjoint union of  $B_2$  and  $A_1 \cap A_2$ , and therefore

$$\mathbb{P}(A_2) = \mathbb{P}(B_2) + \mathbb{P}(A_1 \cap A_2).$$

Since  $\mathbb{P}$  is a nonnegative function, we have  $\mathbb{P}(B_2) \leq \mathbb{P}(A_2)$ . Since  $A_1 \cup A_2 = B_1 \cup B_2$ , we have

$$\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(B_1 \cup B_2) = \mathbb{P}(B_1) + \mathbb{P}(B_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2).$$

□

Two events  $A$  and  $B$  are said to be *independent* if  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ . Events  $A_1, \dots, A_n$  are independent if for any subset  $S \subset [n]$  of the events we have

$$\mathbb{P}\left(\bigcap_{i \in S} A_i\right) = \prod_{i \in S} \mathbb{P}(A_i).$$

If  $A_1, \dots, A_n$  are independent, it implies that any pair  $A_i$  and  $A_j$  are independent, but the converse of this statement is not true (ie, there can be  $A_1, A_2$  and  $A_3$  that are pairwise independent but such that  $A_1, A_2, A_3$  are not independent).

Given a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , a *random variable* is a function  $X : \mathcal{F} \rightarrow \mathbb{R}$  that is measurable (measurable means that for any real number  $t$ , the probability that  $X \leq t$  is well-defined. In our setting  $\Omega$  is finite, and any function  $X$  will be measurable).

## 3.2 The first moment method

In a finite probability space, given a random variable  $X$ , we define the *expected value* of  $X$  to be

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} \mathbb{P}(\omega)X(\omega) = \sum x\mathbb{P}(X = x).$$

(In an infinite probability space, the expected value is defined by an integral). The expected value of a random variable  $X$  is often called the first moment of  $X$ .

### 3.2.1 Linearity of expectation

One very powerful fact is that expectation is linear, ie if  $X$  and  $Y$  are random variables and  $a$  and  $b$  are constants, then

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y).$$

The proof of this follows because integrals are linear (which follows because limits are linear). For example, this means that for random variables  $X_1, \dots, X_n$ , we have

$$\mathbb{E}(X_1 + \dots + X_n) = \mathbb{E}(X_1) + \dots + \mathbb{E}(X_n),$$

and this is true no matter how the  $X_i$ 's are related, even if they are not independent! Let's see how powerful linearity of expectation can be:

The *Ramsey number*  $R(s)$  is defined to be the minimum  $n$  such that any graph on  $n$  vertices contains either a complete graph on  $s$  vertices or an independent set on  $s$  vertices. To show that  $R(s) \leq N$  it means that no matter how I color the edges of  $K_N$  with red and blue, there will be a monochromatic clique on  $s$  vertices. To show that  $R(s) > N$  it means that there exists a graph on  $N$  vertices that does not have either an independent set or a clique on  $s$  vertices.

**Theorem 3.2.1.** *For  $s \geq 3$ , we have  $R(s) > 2^{s/2-1}$ .*

*Proof.* Let us consider a graph on  $n$  vertices where for every pair of vertices we flip a coin and with probability  $1/2$  we put an edge. This determines a probability space (the space of all labeled graphs with the uniform probability distribution). Let  $X$  be the random variable that counts the number of sets of  $s$  vertices such that either all of the  $\binom{s}{2}$  edges are present or none of them are. For any fixed set of  $s$  vertices, the probability that it forms a clique in this random graph is  $(1/2)^{\binom{s}{2}}$  and the probability that it forms an independent set is also  $(1/2)^{\binom{s}{2}}$ .

Given a subset  $S$  of  $s$  vertices, let  $X_S$  denote the number of cliques or independent sets of size  $s$  on those  $s$  vertices (ie  $X_S$  is 1 if  $S$  forms a clique or independent set and 0 otherwise). Then for each  $S$ , we have

$$\mathbb{E}(X_S) = 1 \cdot \mathbb{P}(S \text{ is a clique}) + 1 \cdot \mathbb{P}(S \text{ is an independent set}) + 0 \cdot \mathbb{P}(S \text{ is neither}) = 2^{1-\binom{s}{2}}.$$



Now

$$X = \sum_{S \in \binom{[n]}{s}} X_S$$

and so

$$\mathbb{E}(X) = \binom{n}{s} 2^{1-\binom{s}{2}}.$$

If the expected number of cliques or independent sets in the graph is less than 1, then there must exist a graph that has 0 cliques or independent sets (why?). We have

$$\mathbb{E}(X) = 2 \binom{n}{s} 2^{-\binom{s}{2}} < 2n^s 2^{-s(s-1)/2}.$$

This is less than 1 if

$$2n^s \leq 2^{s(s-1)/2}$$

and so we may take  $n = 2^{(s-1)/2}$ .

□

Indeed, the same proof proves a more general theorem about off-diagonal Ramsey numbers. The Ramsey number  $R(s, t)$  is defined to be the minimum  $n$  such that any red/blue coloring of the edges of  $K_n$  contains either a red  $K_s$  or a blue  $K_t$ . If  $s$  and  $t$  are not equal, we should not flip coins in our probabilistic construction with probability  $1/2$  for each color. Coloring an edge blue with probability  $p$  and red with probability  $1 - p$  and following the proof above yields the following theorem (check the details on your own!).

**Theorem 3.2.2.** *If  $p \in [0, 1]$  satisfies the inequality*

$$\binom{n}{s} (1-p)^{\binom{s}{2}} + \binom{n}{t} p^{\binom{t}{2}} < 1,$$

*Then  $R(s, t) > n$ .*

**Homework.** Use the above theorem to show that  $R(4, t) = \Omega\left(\left(\frac{t}{\log t}\right)^{3/2}\right)$ .

Let's see another application, signing unit vectors:

**Theorem 3.2.3.** *Given any unit vectors  $v_1, \dots, v_n \in \mathbb{R}^n$ , there exists a signing  $\epsilon_i \in \{-1, 1\}$  such that*

$$|\epsilon_1 v_1 + \dots + \epsilon_n v_n| \leq \sqrt{n}.$$

Note that any signing of an orthonormal set of basis vectors in  $\mathbb{R}^n$  will have norm  $\sqrt{n}$ , so this is best possible.

*Proof.* Choose  $\epsilon_i$  independently at random with probability 1/2 for each sign. Let  $v = \sum \epsilon_i v_i$ . Let  $X$  be the random variable defined by  $|v|^2$ . Note that

$$\langle v, v \rangle = \sum_{i,j} \epsilon_i \epsilon_j \langle v_i, v_j \rangle.$$

Note that  $\langle v_i, v_j \rangle$  is a constant. Also note that the signing was chosen independently, so for  $i \neq j$ , we have

$$\mathbb{E}(\epsilon_i \epsilon_j) = \mathbb{E}(\epsilon_i) \mathbb{E}(\epsilon_j) = 0.$$

By linearity of expectation, the expected value of  $X$  is

$$\mathbb{E}(X) = \mathbb{E} \left( \sum_{i,j} \epsilon_i \epsilon_j \langle v_i, v_j \rangle \right) = \sum_{ij} \langle v_i, v_j \rangle \mathbb{E}(\epsilon_i \epsilon_j) = \sum_i |v_i|^2 \mathbb{E}(\epsilon_i^2) = n.$$

Since the expected norm of  $v$  is  $\sqrt{n}$ , there must exist a signing such that the norm of  $v$  is at most  $\sqrt{n}$ . □

### 3.3 Alterations

The most basic version of the first moment method says to construct a combinatorial object randomly and then to show that with positive probability it has the properties that we want. We now consider what to do in the situation where our randomly

constructed object is close to having the desired properties but does not yet. In this situation, we may delete part of or alter our object so that it does have the properties that we want. This is called the alteration method. Let's see two examples.

**Theorem 3.3.1.** *For any  $p \in [0.1]$ , we have*

$$R(s, t) > n - \binom{n}{s} p^{\binom{s}{2}} - \binom{n}{t} (1-p)^{\binom{t}{2}}.$$

*Proof.* Color the edges of  $K_n$  randomly and independently, red with probability  $p$  and blue with probability  $(1-p)$ . Then

$$\binom{n}{s} p^{\binom{s}{2}} + \binom{n}{t} (1-p)^{\binom{t}{2}}$$

is the expected number of red  $K_s$  plus blue  $K_t$ , and so there is a coloring with at most this many. Fix such a coloring. Now, from each red  $K_s$  or blue  $K_t$ , remove one vertex. After this deletion occurs, our graph no longer has any red  $K_s$  or blue  $K_t$  and has at least

$$n - \binom{n}{s} p^{\binom{s}{2}} - \binom{n}{t} (1-p)^{\binom{t}{2}}$$

vertices. □

**Homework.** Use this to show that  $R(4, t) = \Omega\left(\left(\frac{t}{\log t}\right)^2\right)$ .

An *independent set* in a graph  $G$  is a set  $S$  such that  $u, v \in S$  implies  $uv$  is not an edge in  $G$ . The *independence number* of a graph  $G$  is the size of the largest independent set in  $G$  and is denoted by  $\alpha(G)$ .

**Theorem 3.3.2.** *Let  $d \geq 1$  and let  $G$  be a graph with  $\frac{nd}{2}$  edges. Then  $\alpha(G) \geq \frac{n}{2d}$ .*

*Proof.* Choose a set  $S$  randomly by putting each vertex in  $S$  independently with probability  $p$ . We will choose  $p$  later. Let  $X$  be the size of  $S$  and let  $Y$  be the number of edges of  $G$  that have both endpoints in  $S$ . Then the probability that a vertex is

in  $S$  is  $p$  and the probability that an edge is in  $S$  is  $p^2$  (because vertices are put in  $S$  independently).

Given  $S$ , we may choose a subset of  $S$  of size at least  $X - Y$  that is an independent set by removing one vertex from each edge in  $S$ . That is, we take the original  $X$  vertices in  $S$ , we remove at most  $Y$  vertices from  $S$  (one endpoint of each edge), and by removing these vertices we destroy all edges in  $S$ . Therefore

$$\alpha(G) \geq \mathbb{E}(X - Y) = \mathbb{E}(X) - \mathbb{E}(Y) = pn - p^2 \frac{dn}{2}.$$

(note that we have used linearity of expectation multiple times). Choosing  $p = \frac{1}{d}$  yields the result.  $\square$

### 3.4 Markov and Chebyshev

*Chebyshev said it, but I'll say it again: there's always a prime between  $n$  and  $2n$ .*

– Nathan Fine

Next we will develop some tools from probability. The first theorem, Markov's inequality, has a straightforward proof but is useful.

**Theorem 3.4.1** (Markov's Inequality). *Let  $X$  be a nonnegative random variable and  $\lambda > 0$  a real number. Then*

$$\mathbb{P}(X \geq \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

*Proof.* The definition of the expected value of  $X$  gives

$$\begin{aligned} \mathbb{E}(X) &= \sum x \mathbb{P}(X = x) \\ &= \sum_{0 \leq x < \lambda} x \mathbb{P}(X = x) + \sum_{x \geq \lambda} x \mathbb{P}(X = x) \geq \lambda \sum_{x \geq \lambda} \mathbb{P}(X = x) = \lambda \mathbb{P}(X \geq \lambda). \end{aligned}$$

$\square$

We use Markov's inequality to prove a famous theorem of Erdős. The *chromatic number* of a graph  $G$  is the minimum number of colors needed to color the vertices so that no edge has the same color endpoints. The chromatic number of a graph depends on local structure in the graph, eg, if there is a clique of size  $k$  in  $G$  then that clique will need to use  $k$  distinct colors. The following theorem also shows that the chromatic number can depend on the global structure of a graph. The *girth* of a graph  $G$  is the length of its shortest cycle. If a graph has high girth then locally it is very sparse.

**Theorem 3.4.2.** *For every pair of number  $g, k$  there is a graph  $G$  with chromatic number greater than  $k$  and girth greater than  $g$ .*

*Proof.* We take a graph on  $n$  vertices with each edge included independently with probability  $p = \frac{n^\gamma}{n}$  where  $0 < \gamma < 1/g$ . For any  $t$  vertices, the probability that  $(v_1, v_2, \dots, v_t)$  forms a cycle is  $p^t$ . Let  $X$  be the random variable that counts the number of cycles of length at most  $g$  in  $G$ . Then

$$\mathbb{E}(X) = \sum_{3 \leq t \leq g} n(n-1)(n-2) \cdots (n-t+1) \frac{p^t}{2t} \leq gn^{g\gamma} = o(n).$$

Therefore, by Markov's inequality,  $\mathbb{P}(X > n/2) \rightarrow 0$  as  $n \rightarrow \infty$ . Set  $a = \lceil \frac{3}{p} \ln n \rceil$  and let  $Y$  be the random variable that counts the number of independent sets of size  $a$  in  $G$ . Then

$$\mathbb{E}(Y) = \binom{n}{a} (1-p)^{\binom{a}{2}} \leq n^a e^{-p \binom{a}{2}} = \exp(a(\ln n - p(a-1)/2)).$$

This term also goes to 0 as  $n \rightarrow \infty$ . By Markov's inequality

$$\mathbb{P}(\alpha(G) \geq a) = \mathbb{P}(Y \geq 1) \leq \frac{\mathbb{E}(Y)}{1} \rightarrow 0.$$

Therefore, for  $n$  large enough, there is a  $G$  such that the number of cycles of length at most  $g$  is less than  $n/2$  and there is no independent set of size  $a$ . Using the alteration

method, we delete one vertex for each of the cycles to obtain a graph  $G'$  on at least  $n/2$  vertices that has girth  $g$  and independence number at most  $a$ . Then

$$\chi(G') \geq \frac{n/2}{a} = \Omega(n^\gamma).$$

This is larger than  $k$  for  $n$  large enough. □

Sometimes just knowing the expected value of a random variable is not enough. When the first moment fails, try the second moment!

**Definition.** Let  $X$  be a random variable. The *variance* of  $X$  is defined to be

$$\text{Var}[X] = \mathbb{E} [(X - \mathbb{E}[x])^2].$$

Expanding the definition and using linearity of expectation, one sees that

$$\text{Var}[X] = \mathbb{E} [(X - \mathbb{E}[X])^2] = \mathbb{E} [X^2] - (\mathbb{E}[X])^2.$$

The *standard deviation* of a random variable  $X$  is the square root of its variance and is denoted by  $\sigma$ .

One of the most useful properties of expectation is that it is linear, which allowed us to compute the expectation of a sum of random variables by computing their expectations individually, knowing nothing about their possible dependencies. We will not have this luxury with variance. Let  $X$  and  $Y$  be random variables, and let's try to compute the variance of their sum:

$$\begin{aligned} \text{Var}[X + Y] &= \mathbb{E} [(X + Y)^2] - (\mathbb{E}[X + Y])^2 \\ &= (\mathbb{E} [X^2] + 2\mathbb{E}[XY] + \mathbb{E} [Y^2]) - ((\mathbb{E}[X])^2 + 2\mathbb{E}[X]\mathbb{E}[Y] + (\mathbb{E}[Y])^2) \\ &= \text{Var}[X] + \text{Var}[Y] + 2(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]). \end{aligned}$$

This extra term  $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$  is called the *covariance* of  $X$  and  $Y$  and is denoted by  $\text{Cov}[X, Y]$ . You should check that this calculation extends to summing over any number of random variables.

**Theorem.** Let  $X_1, \dots, X_n$  be random variables. Then

$$\text{Var} \left[ \sum_{i=1}^n X_i \right] = \sum_{i=1}^n \text{Var}[X_i] + \sum_{i \neq j} \text{Cov}[X_i, X_j].$$

The variance of a random variable gives us a new tool to obtain concentration results: Chebyshev's Inequality.

**Theorem** (Chebyshev's Inequality). *Let  $X$  be a random variable with  $\text{Var}[X] < \infty$ . Then for any  $\lambda > 0$ , we have*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \lambda] \leq \frac{\text{Var}[X]}{\lambda^2}.$$

*Proof.* Given  $X$  define a random variable  $Y$  by  $Y = (X - \mathbb{E}[X])^2$ . Then  $Y$  is a non-negative random variable, and so by Markov's Inequality we have

$$\mathbb{P}[Y \geq \lambda^2] \leq \frac{\mathbb{E}[Y]}{\lambda^2} = \frac{\text{Var}[X]}{\lambda^2}.$$

Noting that that  $Y \geq \lambda^2$  is equivalent to  $|X - \mathbb{E}[X]| \geq \lambda$  proves the result.  $\square$

We will use Chebyshev's inequality to prove two theorems.

**Theorem 3.4.3.** *Let  $G$  be a graph on  $n$  vertices with each edge included independently with probability  $p$ . Let  $\omega(n)$  be any function that goes to infinity arbitrarily slowly. Then*

*i* *If  $p < \frac{1}{\omega(n) \cdot n}$ ,  $G$  contains a triangle with probability tending to 0.*

ii If  $p > \frac{\omega(n)}{n}$ , then  $G$  contains a triangle with probability tending to 1.

*Proof.* Let  $T$  be the random variable which counts the number of triangles in  $G$  and for each set of 3 vertices, let  $T_1, \dots, T_{\binom{n}{3}}$  be the indicator random variable that that set of vertices forms a triangle. So  $\mathbb{P}(T_i = 1) = p^3$  and  $\mathbb{P}(T_i = 0) = 1 - p^3$ . Then  $T = \sum T_i$ , so  $\mathbb{E}(T) = p^3 \binom{n}{3}$ . If  $p < \frac{1}{\omega(n)n}$ , then  $\mathbb{E}(T) = o(1)$  and so by Markov's inequality,

$$\mathbb{P}(T \geq 1) \leq \frac{\mathbb{E}(T)}{1} = o(1),$$

and the first part is proved. For the second part, we must compute the variance of  $T$ .

$$\text{Var}[T] = \text{Var}\left[\sum T_i\right] = \sum_{i=1}^{\binom{n}{3}} \text{Var}[T_i] + \sum_{i \neq j} \text{Cov}[T_i, T_j].$$

$T_i^2$  is a random variable which takes value 1 with probability  $p^3$  and 0 with probability  $1 - p^3$ . Therefore  $\text{Var}[T_i] = p^3 - p^6$ . For  $T_i$  and  $T_j$ , calculating  $\text{Cov}[T_i, T_j]$  depends on how  $T_i$  and  $T_j$  intersect. If  $T_i$  and  $T_j$  are edge-disjoint, then they are independent random variables and so their covariance is 0. If  $T_i$  and  $T_j$  are not edge-disjoint and  $i \neq j$ , then they must share exactly one edge. In this case,  $T_i T_j$  is a random variable which takes value 1 with probability  $p^5$  and 0 with probability  $1 - p^5$ . So in this case  $\text{Cov}[T_i, T_j] = p^5 - p^6$ . There are less than  $n^4$  pairs  $T_i$  and  $T_j$  which intersect on one edge, so

$$\text{Var}[T] \leq n^3 p^3 + n^4 p^5.$$

By Chebyshev's inequality

$$\mathbb{P}[T = 0] \leq \mathbb{P}[|T - \mathbb{E}[T]| \geq \mathbb{E}[T]] \leq \frac{\text{Var}[T]}{(\mathbb{E}[T])^2} < \frac{p^3 n^3 + p^5 n^4}{p^6 n^6} \leq \frac{\omega(n)^3 + \omega(n)^4/n}{\omega(n)^6} \rightarrow 0.$$

□



For the second application of Chebyshev's inequality, the basic question is this: what is the largest subset  $A$  of  $[n]$  such that all of the subsets of  $A$  have distinct sums. If  $A$  has size  $k$ , then  $A$  has  $2^k$  subsets, all of which must have distinct sums. Further, all of these sums must be less than  $kn$ , and so we must have  $2^k < kn$ , which simplifies to  $k < \log_2 n + \log_2 \log_2 n + 1$ . We use Chebyshev to improve the second term. It is a 500 USD Erdős question to improve the upper bound to  $\log_2 n + O(1)$ .

**Theorem 3.4.4.** *If  $A \subset [n]$  is such that all of its subsets have distinct sums, then  $|A| \leq \log_2 n + \frac{1}{2} \log_2 \log_2 n + 2$ .*

*Proof.* Let  $A = \{x_1, \dots, x_k\} \subset [n]$ . For all  $i$  we will flip a coin and set  $\gamma_i = 0$  with probability  $1/2$  and  $\gamma_i = 1$  with probability  $1/2$ . Define a random variable

$$X = \sum_{i=1}^k \gamma_i x_i.$$

Then  $\mathbb{E}[X] = \frac{1}{2} \sum x_i$  and

$$\text{Var}[X] = \frac{1}{4} \sum_{i=1}^k x_i^2 < \frac{kn^2}{4}.$$

Then by Chebyshev's inequality, we have

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq n\sqrt{k}/\sqrt{2}] \leq \frac{1}{2},$$

which is the same as saying  $\mathbb{P}[|X - \mathbb{E}[X]| < n\sqrt{k}/\sqrt{2}] \geq \frac{1}{2}$ .

Now comes a key step in the proof: the point where we use the assumption that  $A$  has distinct subset sums. Fix any integer  $x$ . Then either  $x$  is the sum of one of the subsets of  $A$  or it is not. If it is not, then  $\mathbb{P}[X = x] = 0$ . If it is, then because the subset sums of  $A$  are distinct, there is a unique choice of  $\gamma$ s such that  $X = x$ , and therefore  $\mathbb{P}[X = x] = 2^{-k}$ . So for any integer  $x$ , the probability that  $X = x$  is either  $2^{-k}$  or 0. Therefore

$$\mathbb{P}[|X - \mathbb{E}[X]| < n\sqrt{k}/\sqrt{2}] = \mathbb{P}[X \text{ lies in a specific set of } \sqrt{2}n\sqrt{k} \text{ integers}] \leq 2^{-k} \sqrt{2}n\sqrt{k},$$

which combined with the above lower bound gives

$$\frac{1}{2} < 2^{-k} \sqrt{2kn}.$$

This is equivalent to the upper bound. □

### 3.5 Chernoff Bound

Chebyshev's inequality gives an upper bound on the probability that a random variable deviates from its mean. Recall that  $\sigma$ , the standard deviation, is the square root of the variance. If  $X$  has standard deviation  $\sigma$ , Chebyshev's inequality says

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq k\sigma] \leq \frac{\text{Var}[X]}{k^2\sigma^2} = \frac{1}{k^2}.$$

So Chebyshev gives us that the probability that a random variable is  $k$  standard deviations away from its mean goes to 0 polynomially in  $k$ . Often we need even better concentration than this. In general we cannot ask for a better bound, but in many situations we can do much better.

Let  $X_1, \dots, X_n$  be independent and identically distributed as  $X_i = -1$  with probability  $1/2$  and  $X_i = 1$  with probability  $1/2$ . Let  $X = X_1 + \dots + X_n$ . Note that the expected value of  $S$  is 0.

**Theorem 3.5.1.** *For any  $k \geq 0$ , we have*

$$\mathbb{P}(X \geq k\sqrt{n}) < e^{-k^2/2} \quad \text{and} \quad \mathbb{P}(X \leq -k\sqrt{n}) < e^{-k^2/2}.$$

Note that the standard deviation of  $S$  is  $\sqrt{n}$ , so this says that the probability of being  $k$  standard deviations away from the mean is exponentially small in  $k$ .

*Proof.* We prove just the upper tail, and the lower tail can be seen by taking complementary events (or symmetry). Let  $t > 0$  be a real number which we will optimize

later. Then

$$\begin{aligned}\mathbb{P}(X \geq k\sqrt{n}) &= \mathbb{P}\left(e^{tX} \geq e^{tk\sqrt{n}}\right) \\ &\leq \frac{\mathbb{E}(e^{tX})}{e^{tk\sqrt{n}}}\end{aligned}$$

where the upper bound is by Markov's inequality. Now  $e^{tX} = e^{tX_1}e^{tX_2} \dots e^{tX_n}$ , and since  $X_1, \dots, X_n$  are independent, we have

$$\mathbb{E}(e^{tX}) = \prod_{i=1}^n \mathbb{E}(e^{tX_i}) = (\mathbb{E}(e^{tX_1}))^n$$

since the  $X_i$ 's are identically distributed. Now  $e^{tX_1}$  takes value  $e^t$  with probability  $1/2$  and  $e^{-t}$  with probability  $1/2$ . Combining, we have

$$\mathbb{P}(X > k\sqrt{n}) \leq \frac{((e^t + e^{-t})/2)^n}{e^{tk\sqrt{n}}} \leq \frac{e^{nt^2/2}}{e^{tk\sqrt{n}}}.$$

The last inequality is using that  $(e^t + e^{-t})/2 \leq e^{t^2/2}$  which can be seen by comparing the Taylor series of each side term by term. Choosing  $t = \frac{k}{\sqrt{n}}$  yields the result.  $\square$

This is called Chernoff's Bound and there are several other slightly messier but more general versions. One general version that is pretty easy to use is the following:

**Theorem 3.5.2.** *Let  $X_1, \dots, X_n$  be independent  $\{0, 1\}$  random variables and let  $X = X_1 + \dots + X_n$ . Let  $\mu = \mathbb{E}(X)$  and let  $\epsilon$  be any real number with  $0 \leq \epsilon \leq 1$ . Then*

$$\mathbb{P}(X \leq (1 - \epsilon)\mu) \leq e^{-\epsilon^2\mu/2} \quad \text{and} \quad \mathbb{P}(X \geq (1 + \epsilon)\mu) \leq e^{-\epsilon^2\mu/3}.$$

Given a set  $v_1, \dots, v_n$  of  $n$  teams, a *tournament* is an assignment for every  $i, j$  that either  $v_i$  beats  $v_j$  or  $v_j$  beats  $v_i$ . A *ranking* of the teams is a permutation  $\sigma \in S_n$  (eg  $\sigma$  maps team  $i$  to their ranking. So if  $\sigma(j) = i$ , then team  $j$  is the best ranked team, and if  $\sigma(k) = n$  then team  $k$  is the worst ranked team). An *upset* given a tournament and a ranking is when  $v_i$  beats  $v_j$  but  $\sigma(j) < \sigma(i)$ . Can we rank any tournament so that there are not very many upsets? Shockingly the answer is no and not by a long shot. We use the Chernoff bound to prove this.

**Theorem 3.5.3** (Erdős and Moon 1963). *Given a tournament on  $n$  vertices, for each  $\sigma \in S_n$  let  $D_\sigma$  be the difference in between the number of upsets and non-upsets with the ranking  $\sigma$ . Then there exists a tournament  $T$  such that for all  $\sigma$ ,  $D_\sigma < n^{3/2}\sqrt{\log n}$ .*

Since the total number of games is  $\binom{n}{2}$ , this means that there are tournaments where we cannot give a ranking that predicts 50.0001 percent of the games correctly.

*Proof.* Let  $\sigma$  be a fixed ranking. In the game between  $v_i$  and  $v_j$  we determine the winner randomly and independently, each with probability  $1/2$ . For each edge  $ij$ , let  $X_{ij} = 1$  if  $\sigma$  correctly predicts the outcome and  $X_{ij} = -1$  if the game is an upset. Then

$$D_\sigma = \sum X_{ij}.$$

$D_\sigma$  has mean 0 and standard deviation  $\sqrt{\binom{n}{2}}$ . Then the Chernoff bound gives that

$$\mathbb{P}\left(|D_\sigma| > k\sqrt{\binom{n}{2}}\right) \leq 2e^{-k^2/2}.$$

Choosing  $k = \sqrt{2n \log n}$  gives

$$\mathbb{P}\left(|D_\sigma| > n^{3/2}\sqrt{\log n}\right) < 2e^{-n \log n} = 2n^{-n}.$$

Since the total number of rankings  $\sigma$  is  $n!$ , the probability that  $D_\sigma$  exceeds  $n^{3/2}\sqrt{\log n}$  for any of them is at most  $n!(2n^{-n}) \rightarrow 0$ .  $\square$

## 3.6 Lovász Local Lemma

We are often trying to show that a certain combinatorial configuration exists by showing that with positive probability, some list of “bad” events does not occur. For example, we showed proved a lower bound for Ramsey numbers by showing that we could color randomly and with positive probability, none of the sets of  $t$  vertices in

the graph formed a monochromatic clique. Let  $A_1, \dots, A_n$  be a set of “bad” events.

We would like to show that

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) > 0.$$

One way to do this is the union bound: If

$$\sum_{i=1}^n \mathbb{P}(A_i) < 1,$$

then with positive probability none of the events occur. Unfortunately, we are not always able to conclude that this sum will be less than 1. We can also get the conclusion if the events are independent, for in that case

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) = (1 - \mathbb{P}(A_1))(1 - \mathbb{P}(A_2)) \cdots (1 - \mathbb{P}(A_n)).$$

Unfortunately, the events we are interested in are usually not independent. The Lovász Local Lemma gives us a way to get to our conclusion if the events are not “too dependent”. Let  $S \subset [n] \setminus \{i\}$ . We say that  $A_i$  is independent of all of the events  $\{A_j : j \in S\}$  if for any subset  $S' \subset S$ , we have

$$\mathbb{P}\left(A_i \cap \bigcap_{j \in S'} A_j\right) = \mathbb{P}(A_i) \cdot \mathbb{P}\left(\bigcap_{j \in S'} A_j\right).$$

For events  $A_1, \dots, A_n$  we define a *dependency graph*  $D = (V, E)$  where  $V(D) = \{A_i\}$  and for each  $i$ ,  $A_i$  is independent of all of the events it is not adjacent to. (Note that this graph is not unique).

**Theorem 3.6.1** (Lovász Local Lemma). *Let  $A_1, \dots, A_n$  be events in a probability space and let  $D$  be a dependency digraph for the  $A_i$  with maximum degree  $d$ . Then if there exists a  $\gamma \in [0, 1)$  such that for all  $i$*

$$\mathbb{P}(A_i) \leq \gamma(1 - \gamma)^d,$$

then

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) \geq (1 - \gamma)^n > 0.$$

The theorem follows from a lemma:

**Lemma 3.6.2.** *For any  $i$ , and any set  $S \in [n] \setminus i$ , we have*

$$\mathbb{P}\left(A_i \mid \bigcap_{j \in S} A_j^c\right) \leq \gamma.$$

*Proof of Theorem.* To prove the theorem from the lemma, note that

$$\begin{aligned} \mathbb{P}\left(\bigcap_{i=1}^n A_i^c\right) &= \mathbb{P}(A_1^c) \mathbb{P}(A_2^c \mid A_1^c) \mathbb{P}(A_3^c \mid A_1^c \cap A_2^c) \cdots \mathbb{P}(A_n^c \mid \bigcap_{i=1}^{n-1} A_i^c) \\ &= \prod_{i=1}^n \mathbb{P}\left(A_i^c \mid \bigcap_{j < i} A_j^c\right) > (1 - \gamma)^n. \end{aligned}$$

□

*Proof of Lemma.* We prove this by induction on  $|S|$ . For the base case  $|S| = 0$ ,  $\mathbb{P}(A_i \mid \bigcap_{j \in S} A_j) = \mathbb{P}(A_i) \leq \gamma(1 - \gamma)^d$  by the hypotheses. This is clearly at most  $\gamma$ . Now we will show it is true for  $|S| = k$  and assume it is true for any  $S'$  with  $|S'| < k$ . Fix  $S$  and let  $I = \{j : j \in S, (k, j) \notin E(D)\}$  and  $D = \{j : j \in S, (k, j) \in E(D)\}$ . ie  $I$  are the set of events in  $S$  that  $A_i$  is independent on and  $D$  is the set of events in  $S$  that  $A_i$  may be dependent on. We make two notes: first that if  $D$  is empty then we are done, so we may assume  $D$  is not empty, and second that

$$\mathbb{P}\left(A_i \mid \bigcap_{j \in S} A_j^c\right) = \mathbb{P}\left(A_i \mid \bigcap_{j \in I \cup D} A_j^c\right).$$

By Bayes' Theorem we have

$$\mathbb{P}\left(A_i \mid \bigcap_{j \in I \cup D} A_j^c\right) = \frac{\mathbb{P}\left((A_i \cap \bigcap_{j \in I} A_j^c) \mid \bigcap_{j \in D} A_j^c\right)}{\mathbb{P}\left(\bigcap_{j \in I} A_j^c \mid \bigcap_{j \in D} A_j^c\right)}.$$

We will upper bound the numerator and lower bound the denominator. For the numerator, we see

$$\mathbb{P}(A_i \cap I | D) \leq \mathbb{P}(A_i | D) = \mathbb{P}(A_i) \leq \gamma(1 - \gamma)^d.$$

For the denominator, let  $I = \{A_{i_1}, \dots, A_{i_t}\}$  (note that  $t < k$ ). We use the induction hypothesis and have

$$\mathbb{P} \left( \bigcap_{j \in I} A_j^c \mid \bigcap_{j \in D} A_j^c \right) = \mathbb{P} \left( A_{i_1}^c \mid \bigcap_{j \in D} A_j^c \right) \mathbb{P} \left( A_{i_2}^c \mid A_{i_1}^c \cap \bigcap_{j \in D} A_j^c \right) \cdots \mathbb{P} \left( A_{i_t}^c \mid \bigcap_{j \in S \setminus \{A_{i_t}\}} A_j^c \right) \geq (1 - \gamma)^d.$$

The last inequality follows since  $D$  has at most  $d$  elements.  $\square$

We give a corollary of this symmetric version as well as the following more general version, the Lopsided Lovász Local Lemma. Both proofs are your homework!

**Corollary 3.6.3.** *Let  $A_1, \dots, A_n$  be events in a probability space and let their dependency graph have maximum degree  $d$ . Assume that there is a  $p$  such that  $\mathbb{P}(A_i) \leq p$  for all  $i$ . Then if  $ep(d + 1) \leq 1$ , we have*

$$\mathbb{P} \left( \bigcap_{i=1}^n A_i^c \right) > 0.$$

**Theorem 3.6.4 (LLLL).** *Let  $A_1, \dots, A_n$  be events in a probability space and let  $D$  be a dependency graph for them. Assume that there exist real number  $x_1, \dots, x_n \in [0, 1)$  such that for all  $i$*

$$\mathbb{P}(A_i) \leq x_i \prod_{(i,j) \in E(D)} (1 - x_j).$$

*Then*

$$\mathbb{P} \left( \bigcap_{i=1}^n A_i^c \right) > 0.$$

We exhibit the power of the local lemma by proving two theorems.

**Theorem 3.6.5.** *Let  $k \geq 9$  and  $G$  be a  $k$ -regular graph. Then there is a vertex coloring of  $G$  with two colors such that every vertex  $v$  has neighbors of each color.*

*Proof.* Color the vertices uniformly and independently, each color with probability  $1/2$ . Let  $A_i$  be the event that vertex  $i$  sees only one color (ie that its neighborhood is monochromatic). Then if any of these events occurs, our coloring is bad, but if all events do not occur simultaneously, then each vertex has a neighbor of each color. We note that

$$\mathbb{P}(A_i) = \frac{1}{2^{k-1}}.$$

Further, each  $A_i$  is dependent only on events from vertices that are at distance 2 from  $i$ . That is, if a vertex  $u$  does not share a neighbor with a vertex  $v$ , then knowing  $A_u$  (or even knowing all of the colors of the neighbors of  $u$ ), does not give us any information about whether  $A_v$  will occur. Since  $G$  is  $k$ -regular, the event  $A_v$  is independent of a set of at least  $|V(G)| - k^2$  events, and so our dependency graph has degree at most  $k^2$ . We can check that for  $k \geq 9$ , we have

$$e \cdot \frac{1}{2^{k-1}} \cdot k^2 < 1$$

and so the local lemma applies to show

$$\mathbb{P} \left( \bigcap_{i \in V(G)} A_i^c \right) > 0.$$

□

**Theorem 3.6.6.** *The Ramsey number  $R(t)$  satisfies*

$$R(t) \geq (1 - o(1)) \frac{\sqrt{2}}{e} t^{t/2}$$

where the  $o(1)$  goes to 0 as  $t \rightarrow \infty$ .



Recall that showing a lower bound  $R(t) \geq m$  means showing that there exists an edge coloring of  $K_m$  such that there is no monochromatic  $K_t$ . This is still the best lower bound that is known.

*Proof.* Color the edges of  $K_n$  randomly where  $n = (1 - o(1)) \frac{\sqrt{2}}{e} t 2^{t/2}$ . Let  $A_1, \dots, A_{\binom{n}{t}}$  be events where we have ordered the  $t$ -subsets of the vertex set in any way and  $A_i$  denotes the event that the  $i$ 'th such set induces a monochromatic  $K_t$ . Then

$$\mathbb{P}(A_i) = 2^{1 - \binom{t}{2}}.$$

Two events  $A_i$  and  $A_j$  are independent unless they share at least one edge. If two distinct cliques share at least one edge, then they share between 2 and  $t - 1$  vertices. Therefore the maximum degree of the dependency graph is

$$\sum_{k=2}^{t-1} \binom{t}{k} \binom{n-t}{t-k}.$$

Using standard estimates on binomial coefficients shows  $ep(d+1) < 1$ . □



# 4

## Extremal Graph Theory

*In this short note, I will restrict myself to Turán's work in graph theory, even though his main work was in analytic number theory and various other branches of real and complex analysis. Turán had the remarkable ability to write perhaps only one paper or to state one problem in various fields distant from his own; later others would pursue his idea and a new subject would be born.*

– Paul Erdős

In extremal graph theory, we are trying to optimize a graph invariant over a fixed family of graphs. This statement is deliberately broad, and an incredible number of interesting problems can be phrased this way. We will mostly focus on two of the most well-studied extremal graph theory problems: Turán-type problems and Ramsey-type problems.

Throughout this chapter,  $G$  will always denote a graph on  $n$  vertices. For a vertex  $v \in V(G)$ ,  $d(v)$  will denote the degree of  $v$ . We will use  $e(G)$  to denote the number of edges in  $G$ .

## 4.1 Turán's Theorem

Let's start with a quantity we have already seen. Given a graph  $F$ , let  $\text{ex}(n, F)$  denote the maximum number of edges in an  $n$  vertex graph which does not contain a copy of  $F$ . Extremal graph theory was born more than a century ago with Mantel's Theorem.

**Theorem 4.1.1** (Mantel's Theorem, 1907). *Let  $G$  be a graph on  $n$  vertices which does not contain a triangle. Then  $e(G) \leq \lfloor \frac{n^2}{4} \rfloor$  with equality if and only if  $G$  is a complete bipartite graph with partite sets of as equal size as possible.*

In our notation, this says that  $\text{ex}(n, K_3) \leq \frac{n^2}{4}$ . As you noticed in lecture, the extremal graph not only forbids  $K_3$  but actually does not contain any odd cycles! This means that  $\text{ex}(n, K_3) = \text{ex}(n, \{K_3, K_5, K_7, \dots\})$ .

*First proof of Mantel's Theorem.* Let  $e(G) = m$ . Note that for any edge  $xy$ , any other vertex can be adjacent to at most one of  $x$  or  $y$  (otherwise we would have a triangle). This means that for any edge  $xy$  we have

$$d(x) + d(y) \leq n.$$

Summing over all of the edges in the graph gives

$$\sum_{x \sim y} d(x) + d(y) \leq mn.$$

On the other hand

$$\sum_{x \sim y} d(x) + d(y) = \sum_{v \in V(G)} (d(v))^2.$$

Now we may use Cauchy-Schwarz to see that

$$\sum_{v \in V(G)} (d(v))^2 \geq \frac{\left(\sum_{v \in V(G)} d(v)\right)^2}{n} = \frac{4m^2}{n}.$$

Rearranging gives  $m \leq \frac{n^2}{4}$ . Go through the proof to see what must happen if we have equality!  $\square$

*Second proof of Mantel's Theorem.* We prove by induction on  $n$ . If  $n = 1$  or  $n = 2$ , then the theorem is trivial, so the base case is done. Now let  $xy$  be an edge in  $G$  and let  $H$  be the graph induced by  $V(G) \setminus \{x, y\}$ . Then

$$e(G) = d(x) + d(y) - 1 + e(H).$$

As we noted in the first proof,  $d(x) + d(y) \leq n$ . By the induction hypothesis,  $e(H) \leq \frac{(n-2)^2}{4}$ . So we have

$$e(G) \leq n - 1 + \frac{(n-2)^2}{4} = \frac{n^2}{4}.$$

$\square$

Turán generalized Mantel's theorem to graphs that do not contain a copy of  $K_r$ . It is easy to see that a complete  $r - 1$  partite graph cannot contain a copy of  $K_r$ : by the pigeonhole principle, for any  $r$  vertices there must be at least 2 in the same partite set, and therefore these two vertices will not have an edge between them, meaning that any set of  $r$  vertices cannot form a  $K_r$ . This gives a lower bound on  $\text{ex}(n, K_r)$  that is roughly  $(1 - \frac{1}{r-1}) \binom{n}{2}$ . Let  $T(n, r - 1)$  be the complete  $(r - 1)$ -partite graph on  $n$  vertices with parts of sizes as equal as possible. Turán's theorem says that no graph that is  $K_r$  free has more edges than  $T(n, r - 1)$ .

**Theorem 4.1.2** (Turán). *Let  $G$  be an  $n$  vertex graph with no copy of  $K_r$  as a subgraph. Then*

$$e(G) \leq e(T(n, r - 1)).$$

There are many proofs of Turán's theorem. This one was told to me by Josh Tobin, and is due to Zykov.

*First proof of Turán's theorem.* Let  $G$  be a graph on  $n$  vertices with no  $K_r$  that has as many edges as possible. That is, assume  $e(G) = \text{ex}(n, K_r)$ .

We define a relation on the vertex set of  $G$ . Let  $xRy$  if  $x$  and  $y$  are *not* adjacent. We claim that this forms an equivalence relation on  $V(G)$ . It is clear that  $xRx$  and that if  $xRy$  and  $yRx$ , so to show that non-adjacency is an equivalence relation we need to show transitivity. (side note: we are only claiming this is an equivalence relation when  $G$  is maximal with respect to not having a  $K_r$ ).

Let  $x, y, z$  be vertices with  $y$  not adjacent to either  $x$  or  $z$ . We assume  $x$  is adjacent to  $z$  and derive a contradiction. If  $d(y) < d(x)$ , then we may make a new graph where we remove  $y$  and make a clone of  $x$ . Note that this cannot create a  $K_r$ , and the new graph will have strictly more edges than  $G$ , contradicting that  $e(G) = \text{ex}(n, K_r)$ . So we must have  $d(y) \geq d(x)$ . Similarly  $d(y) \geq d(z)$ . Now let  $H = G \setminus \{x, y, z\}$ . Note that

$$e(G) = e(H) + d(x) + d(y) + d(z) - 1$$

where the 1 is subtracted because we assumed that  $xz$  is an edge. Now create a new graph  $G'$  by adding 3 copies of the vertex  $y$  to  $H$ . Then

$$e(G') = e(H) + 3d(y) \geq e(H) + d(x) + d(y) + d(z) > e(G).$$

Noting that  $G'$  does not contain a  $K_r$  and has strictly more edges than  $G$  gives us a contradiction.

So we have shown that non-adjacency is an equivalence relation on  $V(G)$ . This means that the vertex set can be partitioned into equivalence classes, ie sets of vertices that are related to each other. This means that  $G$  must be a complete multipartite graph (a vertex must not be adjacent with anything in its equivalence class, and must be adjacent to everything not in its equivalence class). Since  $G$  is  $K_r$ -free, there must be at most  $r - 1$  parts, and to maximize the number of edges we must take  $G = T(n, r - 1)$ . □

*Second Proof of Turán's theorem.* Let  $v_1, \dots, v_n$  be the vertices of  $G$ . Choose a permutation  $\pi \in S_n$  uniformly at random. Create a subset  $S \subset V(G)$  by putting  $v_{\pi(i)}$  in  $S$  if  $v_{\pi(i)}$  is adjacent to all of the members already in  $S$ . Therefore,  $S$  induces a clique, and so at the end of the process  $|S| \leq r - 1$ . Let  $X_i$  be the indicator random variable that  $v_i \in S$ , so  $|S| = \sum X_i$ . Then

$$r - 1 \geq \mathbb{E}(|S|) = \sum_{i=1}^n \mathbb{P}(X_i) \geq \sum_{i=1}^n \frac{1}{n - d(v_i)}.$$

The last inequality is because a vertex  $v_i$  to be in  $S$  if it comes before all of its  $n - d(v_i) - 1$  non-neighbors in the permutation  $\pi$ . Now we use Cauchy-Schwarz with  $u_i = \sqrt{n - d(v_i)}$  and  $v_i = \frac{1}{\sqrt{n - d(v_i)}}$ , which gives

$$n^2 \leq \left( \sum_{i=1}^n n - d(v_i) \right) \left( \sum_{i=1}^n \frac{1}{n - d(v_i)} \right) \leq (r - 1) \left( \sum_{i=1}^n n - d(v_i) \right).$$

Noting that  $\sum d(v_i) = 2e(G)$  and rearranging gives

$$e(G) \leq \frac{(r - 2)}{2(r - 1)} n^2.$$

□

So this is very good! When the graph we are forbidding is complete, Turán's theorem gives us an exact result. Exact results are rare in extremal graph theory, and so we are happy. What can we say more generally? What is  $\text{ex}(n, F)$  if  $F$  is not a complete graph? If we are willing to give up a small error term, in many cases we can say something surprisingly sharp.

Let's consider lower bounds first. The Turán graph with  $r$  parts was a good candidate to exclude a complete graph on  $r + 1$  vertices because of the pigeonhole principle: for any  $r + 1$  vertices, there must be a part with at least 2 of them, and so these 2 vertices cannot be adjacent. But if we think a little bit we realize that the

Turán graph is a good graph to exclude many other graphs as well. Say  $\chi(F) = r + 1$ . Then we claim that  $F$  cannot be a subgraph of  $T(n, r)$ . Why?  $T(n, r)$  is a complete  $r$ -partite graph. In particular, coloring each partite set with its own color gives a proper  $r$ -coloring of  $T(n, r)$ . This means that for any graph  $H$  which is a subgraph of  $T(n, r)$ , there is a proper coloring of  $H$  with  $r$  colors. Therefore, since  $F$  has chromatic number strictly greater than  $r$ , it cannot be a subgraph of  $T(n, r)$ . So when  $\chi(F) = r + 1$ , we have shown

$$\text{ex}(n, F) \geq e(T(n, r)) \sim \left(1 - \frac{1}{r}\right) \binom{n}{2}.$$

Is this lower bound best possible? The celebrated Erdős-Stone theorem says, up to a small error, yes.

**Theorem 4.1.3** (Erdős-Stone theorem). *Let  $F$  be a graph with  $\chi(F) = r + 1$ . Then*

$$\text{ex}(n, F) = \left(1 - \frac{1}{r} + o(1)\right) \binom{n}{2}$$

where  $o(1)$  goes to 0 as  $n$  goes to infinity.

We already certified the lower bound. In order to prove the upper bound, one has to show that for any  $\epsilon > 0$ , there is an  $N$  such that for  $n \geq N$ , any graph with at least  $(1 - \frac{1}{r} + \epsilon) \binom{n}{2}$  edges contains a copy of  $F$ . The proof outline is roughly this:

1. Assume  $G$  is a graph with at least  $(1 - \frac{1}{r} + \epsilon) \binom{n}{2}$  edges. Our goal is to show that  $F$  is a subgraph of  $G$ .
2. By removing vertices of small degree, show that it suffices to consider the case where  $G$  has minimum degree at least  $(1 - \frac{1}{r} + \epsilon)n$ . This part is the same idea as number 1 on Homework 7!
3. Show that for any constant  $C$  that we choose,  $G$  must contain a complete  $(r + 1)$ -partite graph with  $C$  vertices in each part. This is the key step in the proof of



the theorem. The proof is by induction on  $r$  and is the same double counting and convexity technique that you will use to prove the Kővari-Sós-Turán theorem on Homework 7.

4. Show that this implies the theorem. That is, once we have a complete  $(r + 1)$ -partite subgraph of  $G$ , we can take  $C$  to be large enough to embed our copy of  $F$  (eg if we take  $C > |V(F)|$  then we can certainly find a copy of  $F$ ).

So this is a very satisfying result. In most cases, this gives us an asymptotic formula for  $\text{ex}(n, F)$ . However, if  $F$  is bipartite, then the theorem simply says that  $\text{ex}(n, F) = o(n^2)$ . In general, finding the Turán number for a bipartite graph is a very difficult problem. For most bipartite graphs, not even the order of magnitude is known. The following theorem, which you will prove in your homework, gives a better upper bound for  $\text{ex}(n, F)$  when  $F$  is bipartite.

**Theorem 4.1.4** (Kővari-Sós-Turán theorem). *Let  $t \geq s \geq 2$ . Then there is a constant  $c$  such that*

$$\text{ex}(n, K_{s,t}) \leq cn^{2-1/s}.$$

Since any bipartite graph  $F$  is a subgraph of  $K_{s',t'}$  for some constants  $s'$  and  $t'$ , we have that  $\text{ex}(n, F) \leq \text{ex}(n, K_{s',t'}) = O(n^{2-1/s'})$ .

We will prove the theorem when  $s = 2$ . You will give the proof of the full theorem in your homework.

*Proof of KST theorem for  $s = 2$ .* Let  $G$  be an  $n$  vertex graph that has no copy of  $K_{2,t}$ . Then since  $G$  is  $K_{2,t}$  free, it means that for any vertices  $x$  and  $y$ , they can have at most  $t - 1$  common neighbors. We write  $d(x, y)$  to denote  $|\Gamma(x) \cap \Gamma(y)|$ , the number of common neighbors of  $x$  and  $y$ . So

$$\sum_{x,y} d(x, y) \leq (t - 1) \binom{n}{2}.$$

On the other hand

$$\sum_{x,y \in V(G)} d(x,y) = \sum_{v \in V(G)} \binom{d(v)}{2} = \frac{1}{2} \sum ((d(v))^2 - d(v)) = \left( \frac{1}{2} \sum (d(v))^2 \right) - e(G).$$

Now by Cauchy-Schwarz, we have

$$\sum (d(v))^2 \geq \frac{(\sum d(v))^2}{n} = \frac{4(e(G))^2}{n}.$$

Combining, this gives

$$\frac{2(e(G))^2}{n} - e(G) \leq (t-1) \binom{n}{2}.$$

Using the quadratic formula to solve for  $e(G)$  gives

$$e(G) \leq \frac{1 + \sqrt{1 + 4(t-1) \binom{n}{2} \frac{2}{n}}}{\frac{2}{n}} = \frac{n + n\sqrt{1 + 4(n-1)(t-1)}}{2} \sim \sqrt{t-1} n^{3/2}.$$

□

The best lower bound we have seen for  $\text{ex}(n, K_{2,2}) = \text{ex}(n, C_4)$  is  $\Omega(n^{4/3})$ . Which is right? We will see shortly.

## 4.2 Projective planes

*The survival of finite geometry as an active field of study depends on someone finding a finite projective plane of a non-prime-power order.*

– Gary Ebert

In this section we foray into the field of finite incidence geometry. A point-line incidence structure (sometimes also called: a rank 2 geometry) is a set  $\mathcal{P}$  of *points* and a set  $\mathcal{L}$  of *lines*, where each line is a subset of the point set. We say that a point  $p$  is *incident* with a line  $l$  if  $p \in l$ . We say the the point  $p$  is on  $l$  to mean the same thing.

**Definition 4.2.1.** A *projective plane* is a point line incidence structure that satisfies the following 3 axioms:

1. For every pair of points, there is a unique line which they are both on.
2. For every pair of lines, they intersect in a unique point.
3. The plane is non-degenerate.

There are two degenerate planes which trivially satisfy the first 2 axioms: if you have only one line in your geometry and all of the points are on it, or if you have only one point in your geometry and all of the lines go through it. It turns out that besides these two examples, the first 2 axioms determine many structural properties that a projective plane must have.

**Theorem 4.2.2.** *Let  $\Pi$  be a finite projective plane. Then there is an integer  $q$  such that every point is on  $q + 1$  lines and every line contains  $q + 1$  points.*

*Proof.* Let  $p$  be a point and  $l$  be a line which does not contain  $p$ . Then for every point  $q \in l$ , the first axiom tells us that there is a unique line  $l_{pq}$  that goes through both  $p$  and  $q$ . Axiom 2 tells us that for every line going through  $p$ , it meets  $l$  in a unique point. Thus there is a bijection from the set of lines going through  $p$  to the set of points on  $l$ , and therefore the sizes of these sets must be the same. But  $p$  and  $l$  were arbitrary. Choosing different pairs of points and lines shows that all points have the same number of lines going through them and all lines have the same number of points on them, and that these numbers are the same. Call this number  $q + 1$ .  $\square$

We call  $q$  the *order* of the projective plane.

**Theorem 4.2.3.** *If  $\Pi$  is a projective plane of order  $q$ , then it has  $q^2 + q + 1$  points and  $q^2 + q + 1$  lines.*

*Proof.* Let  $p$  be a point. Consider the  $q + 1$  lines that go through  $p$ . These lines all meet at  $p$ , but besides this are disjoint (by axiom 2). Furthermore, any point in the projective plane must be on one of these lines. To see this, axiom 1 says that for any point  $q$ , there is a unique line going through  $p$  and  $q$ , and so  $q$  must be on one of the lines through  $p$ . Therefore, we can count the number of points in the projective plane to be:

$$1 + (q + 1)(q + 1 - 1) = q^2 + q + 1.$$

The proof for lines can be done similarly, or one can double count to show that the number of points must be the same as the number of lines.  $\square$

We now show that a projective plane of order  $q$  exists whenever  $q$  is a prime power. It is the biggest open problem in finite geometry to determine whether or not this is the only time when a plane exists.

**Theorem 4.2.4.** *Let  $q$  be a prime power. Then there is a projective plane of order  $q$ .*

*Proof.* Let  $\mathbb{F}_q$  be the finite field of order  $q$ . Let  $V$  be a 3-dimensional vector space over  $\mathbb{F}_q$ . We define an incidence structure as follows:

- Let  $\mathcal{P}$  be the set of 1-dimensional subspaces of  $V$ .
- Let  $\mathcal{L}$  be the set of 2-dimensional subspaces of  $V$ .
- Define incidence by containment. ie a point (1-d subspace) is on a line (2-d subspace) if the 1-d subspace is contained in the 2-d subspace.

Since two distinct 1-dimensional subspaces span a unique 2-dimensional subspace, this structure satisfies the first axiom. Since  $V$  is 3-dimensional, any pair of 2-dimensional subspaces must meet in a 1-dimensional subspace, so the second axiom is also satisfied. It is easy to see that the plane is not degenerate.  $\square$

### 4.3 Sidon sets

*At the very beginning, there was Simon Sidon.*

– Alain Plagne

Recall from your second exam that if  $\Gamma$  is an abelian group, and  $A \subset \Gamma$  is called a *Sidon set* if it has only trivial solutions to the Sidon equation

$$a + b = c + d.$$

That is, if  $a, b, c, d \in A$  and  $a + b = c + d$ , then it implies that  $\{a, b\} = \{c, d\}$ . Note that if  $A$  is a Sidon set, then all of the sums are distinct, but also all of the nonzero differences are distinct. This implies that if  $A \subset \Gamma$  is a Sidon set, then

$$|A|(|A| - 1) \leq |\Gamma| - 1.$$

Finding the maximum size of a Sidon set in  $[n]$  is a 500 USD Erdős problem:

Denote by  $f(n)$  the largest integer  $k$  for which there is a sequence  $1 \leq a_1 < \dots < a_k \leq n$  so that all the sums  $a_i + a_j$  are distinct. Turán and I conjectured about 40 years ago [5] that

$$f(n) = n^{1/2} + O(1). \tag{1}$$

The conjecture seems to be very deep and I offered long ago a prize of 500 dollars for a proof or disproof of (1). The sharpest known results in the direction of (1) state [5]

$$n^{1/2} - n^{1/2-c} < f(n) < n^{1/2} + n^{1/4} + 1. \tag{2}$$

We show that there are Sidon sets of size asymptotic to the upper bound:

**Theorem 4.3.1.** *Let  $p$  be an odd prime. There is a Sidon subset of  $\mathbb{F}_p \times \mathbb{F}_p$  of size  $p$ .*

*Proof.* Let

$$A = \{(x, x^2) : x \in \mathbb{F}_p\}.$$

It is clear that  $|A| = p$ . We now show that  $A$  is a Sidon set. Assume that

$$(a, a^2) + (b, b^2) = (c, c^2) + (d, d^2)$$

with  $a, b, c, d \in \mathbb{F}_p$ . We must show that  $\{a, b\} = \{c, d\}$ . Without loss of generality, assume that  $a \neq c$ . Then we have the following system of equations

$$\begin{aligned} a + b &= c + d \\ a^2 + b^2 &= c^2 + d^2 \end{aligned}$$

which is equivalent to

$$\begin{aligned} a - c &= d - b \\ a^2 - c^2 &= d^2 - b^2. \end{aligned}$$

Since  $a \neq c$  and we are working in a field, we may divide through the second equation by  $a - c = d - b$  to find that

$$a + c = d + b.$$

But this along with  $a - c = d - b$  implies that  $2a = 2d$  which means that  $a = d$  (since  $p$  is odd), and  $b = c$ . □

## 4.4 Constructing $C_4$ -free graphs

Why did we talk about projective planes and Sidon sets? It turns out one can use them to construct  $C_4$ -free graphs.

Let  $\Pi$  be a projective plane of order  $q$ . We will make the bipartite incidence graph of  $\Pi$  as follows (also called the Levi graph): Let  $\mathcal{P}$  and  $\mathcal{L}$  be the points and lines of  $\Pi$ . We define our graph  $G$  by  $V(G) = \mathcal{P} \cup \mathcal{L}$  and for  $p \in \mathcal{P}$ ,  $l \in \mathcal{L}$  we have  $pl \in E(G)$  if and only if  $p \in l$  in  $\Pi$ . Why is this graph  $C_4$ -free? We must show that for any pair

of vertices, they have at most 1 neighbor in common. If one vertex is a point, and one vertex is a line, then the pair have no neighbors in common since  $G$  is bipartite. Now, if I have two vertices which are points, how many neighbors do they have in common? The neighborhood of a point in  $G$  is the set of lines that it is on in  $\Pi$ . By the axioms of a projective plane, a pair of points are on a unique common line. This means that in  $G$ , any pair of points have exactly one neighbor in common. Similarly, any pair of lines meet in a unique point. In  $G$ , the neighborhood of a line is the set of points on it. Therefore, any two lines in  $G$  have a unique common neighbor. So  $G$  is a  $q + 1$  regular graph on  $2(q^2 + 2 + 1)$  points which is  $C_4$ -free. Since we showed that projective planes of order  $q$  exist if  $q$  is a prime power, we have the following corollary:

**Corollary 4.4.1.** *If  $q$  is a prime power, then*

$$\text{ex}(2(q^2 + q + 1), C_4) \geq (q + 1)(q^2 + q + 1).$$

Using a density of primes argument that is the same as in your homework, this implies that

$$\text{ex}(n, C_4) \gtrsim \frac{1}{2\sqrt{2}} n^{3/2}.$$

So we have determined the order of magnitude of  $\text{ex}(n, C_4)$ . In your homework, you will show that you can remove the constant multiple of  $\sqrt{2}$  difference in the upper and lower bounds by smashing together the two partite sets of the above graph in a nice way.

Next we use Sidon sets to show an alternate way to get rid of the  $\sqrt{2}$  factor. Given an abelian group  $\Gamma$  and a set  $A \subset \Gamma$ , we can construct a *Cayley sum graph*. Let  $V(G) = \Gamma$ , and let  $x \sim y$  in  $G$  if and only if  $x + y \in A$ . Note that this graph may have loops (if  $x + x \in A$ ), but if we remove the loops then the graph has minimum degree  $|A| - 1$ . We claim that if  $A$  is a Sidon set, this graph is  $C_4$ -free. To see this, consider

a purported  $C_4$ . That is, let  $x, y, w, z \in \Gamma$  and assume  $xy, yw, wz, zx \in E(G)$ . Then there are  $a, b, c, d \in A$  such that

$$x + y = a$$

$$y + w = b$$

$$w + z = c$$

$$z + x = d$$

But then we have

$$x + y + w + z = a + c = b + d,$$

where  $a, b, c, d \in A$ . Since  $A$  is a Sidon set, we must have  $\{a, c\} = \{b, d\}$ . If  $a = b$  then  $x = w$  and if  $a = d$  then  $y = z$ . In either case,  $xywz$  is not a  $C_4$  in the graph. By the construction of a Sidon set of size  $p$  in the group  $\mathbb{F}_p \times \mathbb{F}_p$ , we have constructed a graph on  $p^2$  vertices with minimum degree at least  $p - 1$ . Therefore we have, for  $p$  a prime:

$$\text{ex}(p^2, C_4) \geq \frac{1}{2}p^2(p - 1).$$

By a density of prime numbers argument, we have the asymptotic formula  $\text{ex}(n, C_4) \sim \frac{1}{2}n^{3/2}$ .

## 4.5 Ramsey numbers

Recall that given graphs  $G$  and  $H$ , the Ramsey number  $R(G, H)$  is the minimum  $n$  such that any red/blue coloring of the edges of  $K_n$  contains a copy of  $G$  with all red edges or a copy of  $H$  with all blue edges. We write  $R(s, t)$  for  $R(K_s, K_t)$ . Giving a lower bound that  $R(s, t) > m$  means coloring the edges of  $K_m$  such that there is no red  $K_s$  or blue  $K_t$ . The first construction one might think of is to take  $t - 1$  disjoint



copies of  $K_{s-1}$  which are colored red, and color the edges in between blue. This gives

$$R(s, t) > (s - 1)(t - 1).$$

For a while, it was thought that maybe this was the right order of magnitude. We saw using the probabilistic method that one can do much, much better. The first moment method gives

$$R(t, t) > 2^{t/2-1}$$

and the Lovász Local Lemma gives

$$R(t, t) \geq (1 - o(1)) \frac{\sqrt{2}s}{e} 2^{s/2}.$$

This is still the best general lower bound. Let's turn to upper bounds.

**Theorem 4.5.1.** *Let  $s, t \geq 2$ . Then*

$$R(s, t) \leq \binom{s+t-2}{t-1}.$$

*Proof.* We prove this by induction. The statement is clear for  $R(s, 2)$  and  $R(2, t)$ . Now consider a red-blue coloring of the edges of  $K_n$  and assume there is no red  $K_s$  or blue  $K_t$ . We must show that  $n < \binom{s+t-2}{t-1}$ . Consider a vertex  $v$  and consider its red neighborhood and its blue neighborhood. By induction, if the number of red edges incident with  $v$  is at least  $\binom{(s-1)+t-2}{t-1}$ , then in the graph induced by its red neighborhood contains either a red  $K_{s-1}$  or a blue  $K_t$ , and so the complete graph has either a red  $K_s$  or a blue  $K_t$ . So the red neighborhood may have size at most  $\binom{s+t-3}{t-1} - 1$ . Similarly, the blue neighborhood of  $v$  may have size at most  $\binom{s+t-3}{t-2} - 1$ . Therefore the number of vertices in the graph is at most

$$1 + \left( \binom{s+t-3}{t-1} - 1 \right) + \left( \binom{s+t-3}{t-2} - 1 \right) = \binom{s+t-2}{t-1} - 1.$$

□

This gives as a corollary that

$$R(t, t) = O\left(\frac{4^t}{t}\right).$$

The best known bound is due to David Conlon, who showed

$$R(t, t) \leq t^{-\Omega\left(\frac{\log t}{\log \log t}\right)} 4^t.$$

We would also like to define multicolor Ramsey numbers. Define

$$R_k(s_1, \dots, s_k)$$

to be the minimum  $n$  such that any  $k$  coloring of the edge set of  $K_n$  contains a monochromatic copy of  $K_{s_i}$  in color  $i$ . A priori, it is not even clear that these numbers should be finite!

**Theorem 4.5.2.** *For any integers  $s_1, \dots, s_k$ ,*

$$R_k(s_1, \dots, s_k) < \infty.$$

*Proof.* We prove by induction on  $k$ . We have already proved the base case when  $k = 2$ . Now let  $K_n$  be colored with  $k$  colors. Given this coloring we create a red/blue coloring of  $K_n$  where we color an edge red if it was colored with color 1, and we color an edge blue if it was colored with colors between 2 and  $k$ . Note that by the induction hypothesis  $R_{k-1}(s_2, \dots, s_k)$  is a finite number. So if  $n = R(s_1, R_{k-1}(s_2, \dots, s_k))$ , then by the base case there exists either a red  $K_{s_1}$  or a blue  $K_{R_{k-1}(s_2, \dots, s_k)}$ . If there is a red  $K_{s_1}$  that means there is a clique of size  $s_1$  in color 1 and we are done. If there is a blue  $K_{R_{k-1}(s_2, \dots, s_k)}$  then there is a clique on  $R_{k-1}(s_2, \dots, s_k)$  vertices using only colors  $2, \dots, k$ . By the induction hypothesis, this clique contains a clique of size  $s_i$  in color  $i$  for some  $i$  between 2 and  $k$ .  $\square$

## 4.6 Combinatorial Number Theory

Ramsey-type theorems do not just have to be on graphs:

**Theorem 4.6.1** (Schur's Theorem). *For any  $k$ , there exists an integer  $n$  such that for any  $k$  coloring of the integers  $\{1, \dots, n\}$ , there are three integers  $x, y, z \in [n]$  with the same color such that  $x + y = z$ .*

*Proof.* Choose  $n = R_k(3, 3, \dots, 3)$ , and assume  $[n]$  is colored with  $k$  colors by the function  $\chi : [n] \rightarrow [k]$ . We define an edge coloring of  $K_n$  by coloring the edge  $ij$  with the color  $\chi(|i - j|)$ . Then by the graph Ramsey theorem, there must be a monochromatic triangle in this graph. Consider a monochromatic triangle  $ijk$  in  $K_n$  and assume  $i < j < k$ . Then we have

$$\chi(j - i) = \chi(k - i) = \chi(k - j).$$

Letting  $x = j - i$ ,  $y = k - j$  and  $z = k - i$  gives three integers in  $[n]$  with the same color such that  $x + y = z$ . □

We use Schur's Theorem to prove that Fermat's Last Theorem does not hold over finite fields:

**Theorem 4.6.2.** *For each  $n$ , there exists a  $p_0$  such that for any prime  $p \geq p_0$  the equation*

$$x^n + y^n \equiv z^n \pmod{p}$$

*has a solution.*

*Proof.* Note that the multiplicative group  $(\mathbb{Z}/p\mathbb{Z})^*$  is cyclic, and so there is a generator  $g$  such that  $(\mathbb{Z}/p\mathbb{Z})^* = \{g^0, g^1, g^2, \dots, g^{p-2}\}$ . For each  $x \in (\mathbb{Z}/p\mathbb{Z})^*$  define  $k_x$  and  $i_x$  so that

$$x = g^{k_x n + i_x}.$$

Now we define a coloring of the integers  $c : \{1, \dots, p\} \rightarrow [n]$  where  $c(x) = i_x$  (note that I am thinking of  $x$  as both an integer as an element of the group simultaneously). For  $p$  large enough, Schur's theorem guarantees a monochromatic solution to  $x' + y' = z'$ . By the way that we have defined our coloring, this means that there exists an  $i$  so that

$$g^{k_{x'}n+i} + g^{k_{y'}n+i} \equiv g^{k_{z'}n+i} \pmod{p}.$$

Dividing through by  $g^i$  and taking  $x = g^{k_{x'}}$ ,  $y = g^{k_{y'}}$  and  $z = g^{k_{z'}}$  gives the result.  $\square$

We now discuss van der Waerden numbers.

**Definition 4.6.3.** Define  $w(r, k)$  to be the minimum  $n$  such that any  $r$  coloring of the integers  $\{1, \dots, n\}$  contains a monochromatic  $k$  term arithmetic progression.

The coloring

12345678

shows that  $w(2, 3) > 8$ . We will now show that  $w(2, 3) \leq 325$ . Assume the integers  $[325]$  are colored red and blue. Partition  $[325]$  into 65 blocks  $B_1, \dots, B_{65}$  as

$$\{1, 2, 3, 4, 5\}\{6, 7, 8, 9, 10\} \dots \{321, 322, 323, 324, 325\}.$$

Since there are 32 ways to color a block of 5 integers, there are two blocks in  $B_1, \dots, B_{33}$  that have the same coloring. That is, there is a  $0 \leq b_i < b_{i+d} \leq 32$  such that  $5b_i + k$  and  $5b_{i+d} + k$  have the same color for  $k \in \{1, \dots, 5\}$ . By the pigeonhole principle again, at least 2 of  $5b_i + 1, 5b_i + 2, 5b_i + 3$  have the same color. Without loss of generality assume that the color is red and that the elements are  $a$  and  $a + e$  (where  $e \in \{1, 2\}$ ). Then we know that  $\{a, a + e, a + 5d, a + 5d + e\}$  are all colored red.  $a + 2e$  or  $a + 5d + 2e$  is red, then we have a red 3 term AP and we are done, so assume that both of these are blue.

But then what color is  $a + 10d + 2e$ ? If it is red then  $\{a, a + 5d + e, a + 10d + 2e\}$  are red and if it is blue then  $\{a + 2e, a + 5d + 2e, a + 10d + 2e\}$  are blue. In general, Gowers has the best upper bound for  $w(r, k)$  with

$$w(r, k) \leq 2^{2^r 2^{k+9}}.$$

Graham conjectures (and offers 1000 USD for)

$$w(2, k) \leq 2^{k^2}.$$

The Lovász Local Lemma gives

$$w(r, k) \geq (1 - o(1)) \frac{r^{k-1}}{4k} \left(1 - \frac{1}{k}\right),$$

and Szabó improved this slightly. Berlekamp showed that for  $p$  a prime

$$w(2, p + 1) > 2^{p+1},$$

and Blankenship, Cummings, and Taranchuk generalized this to colorings with  $r$  colors.

Here we give an upper bound for  $W(r, k)$  that is worse than even a tower.

**Theorem 4.6.4.** *For all  $r, k$ ,  $w(r, k) < \infty$ .*

*Proof.* We prove by double induction on  $r$  and  $k$ . The base cases  $w(1, k) = k$  for all  $k$  and  $w(r, 2) = r + 1$  for all  $r$  are easy. Next we will assume  $w(r, k - 1)$  exists for all  $r$  and show that  $w(r, k)$  exists for all  $r$ .

We define a *sunflower* with  $m$  petals of length  $k - 1$  to be a collection of  $m$

arithmetic progressions of length  $k - 1$ ,

$$\begin{array}{ccc}
 a + d_1 & a + 2d_1 & \cdots & a + (k - 1)d_1 \\
 a + d_2 & a + 2d_2 & \cdots & a + (k - 1)d_2 \\
 & \vdots & & \vdots \\
 a + d_m & a + 2d_m & \cdots & a + (k - 1)d_m
 \end{array}$$

where each AP is monochromatic but different APs have distinct colors. Note that if there is a sunflower with  $m$  petals of length  $k - 1$ , then if  $a$  is colored with any of those  $m$  colors, we have created a monochromatic  $k$  term AP. The theorem then follows after the following lemma.

**Lemma 4.6.5.** *Suppose  $w(r, k - 1)$  exists for all  $r$ . Then for any  $m$ , there exists an  $n = w(r, m, k - 1)$  such that any  $r$  coloring of  $[n]$  contains either a monochromatic  $k$ -term AP or a sunflower with  $m$  petals of length  $k - 1$ .*

We prove this by induction on  $m$ . The base case follows by the assumption that  $w(r, k - 1) < \infty$ . Now assume the statement is true for  $m - 1$  and let  $n_1 = w(r, m - 1, k - 1)$  and  $n_2 = 2w(r^{n_1}, k - 1)$ . Color  $[n_1 n_2]$  and consider  $[n_1 n_2]$  as split into  $n_2$  blocks of length  $n_1$ . Since there are only  $r^{n_1}$  possible colorings of each block, then by the definition of  $n_2$ , there is a  $j$  and a  $d$  with  $n_2/2 \leq j < j + (k - 1)d$  such that all of the blocks  $B_{j+id}$  have the same coloring for  $0 \leq i \leq k - 1$ . Let  $c = dn_1$ .

By the induction hypothesis and the definition of  $n_1$ , each of these blocks contains either a  $k$ -term AP or a sunflower with  $m - 1$  petals of length  $k - 1$ . If there is a  $k$ -term AP we are done, so assume the latter. Since the  $k - 1$  blocks are colored identically, we have for each color  $k - 1$  APs each of length  $k - 1$  which are “evenly spaced”. That is, in color 1 we see

$$\begin{array}{llll}
a + d_1 & a + 2d_1 & \cdots & a + (k - 1)d_1 \\
a + d_1 + c & a + 2d_1 + c & \cdots & a + (k - 1)d_1 + c \\
& \vdots & & \vdots \\
a + d_1 + (k - 1)c & a + 2d_1 + (k - 1)c & \cdots & a + (k - 1)d_1 + (k - 1)c
\end{array}$$

In color 2 we see

$$\begin{array}{llll}
a + d_2 & a + 2d_2 & \cdots & a + (k - 1)d_2 \\
a + d_2 + c & a + 2d_2 + c & \cdots & a + (k - 1)d_2 + c \\
& \vdots & & \vdots \\
a + d_2 + (k - 1)c & a + 2d_2 + (k - 1)c & \cdots & a + (k - 1)d_2 + (k - 1)c
\end{array}$$

etc, with color  $m - 1$  seeing

$$\begin{array}{llll}
a + d_{m-1} & a + 2d_{m-1} & \cdots & a + (k - 1)d_{m-1} \\
a + d_{m-1} + c & a + 2d_{m-1} + c & \cdots & a + (k - 1)d_{m-1} + c \\
& \vdots & & \vdots \\
a + d_{m-1} + (k - 1)c & a + 2d_{m-1} + (k - 1)c & \cdots & a + (k - 1)d_{m-1} + (k - 1)c
\end{array}$$

Now if  $a$  has any of the first  $m - 1$  colors, then we have a  $k$  term AP. If  $a$  has a different color, then there is a sunflower with  $m$  petals of length  $k - 1$  centered at  $a + (k - 1)c$ . □

## 4.7 Erdős-Ko-Rado Theorem

Let  $2^{[n]}$  denote the set of subsets of  $[n]$ . A family of subsets  $\mathcal{F} \subset 2^{[n]}$  is called *intersecting* if for all  $F, F' \in \mathcal{F}$ ,  $F \cap F' \neq \emptyset$ . The subset of  $2^{[n]}$  of subsets which contain  $n$  is an intersecting (each pair intersects at at least the integer  $n$ ) family of size  $2^{n-1}$ . If  $n$  is odd, then the set of subsets of size at least  $n/2$  is also an intersecting family of size  $2^{n-1}$ .

**Theorem 4.7.1.** *If  $\mathcal{F}$  is an intersecting family, then  $|\mathcal{F}| \leq 2^{n-1}$ .*

*Proof.* If  $\mathcal{F}$  is an intersecting family and  $F \in \mathcal{F}$ , then  $F^c$  is not in  $\mathcal{F}$ . Therefore, there is an injection from sets in  $\mathcal{F}$  to sets not in  $\mathcal{F}$ , which means that  $\mathcal{F}$  can have at most half of the possible subsets of  $[n]$ .  $\square$

What if we restrict our sets to have a fixed size? This is the famous Erdős-Ko-Rado Theorem.

**Theorem 4.7.2** (Erdős-Ko-Rado 1961). *Fix a natural number  $k$ . If  $n \geq 2k$ , and  $\mathcal{F}$  is an intersecting family where each element of  $\mathcal{F}$  has size  $k$ , then*

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

Choosing every  $k$  element subset which contains a fixed element shows that this is best possible. Furthermore, if  $n < 2k$ , then one may take every subset of size  $k$  and the family will be intersecting.

*Proof.* Choose  $\pi \in S_n$  uniformly at random. Place the integers  $[n]$  in a circle in the order prescribed by  $\pi$ . For a set  $F \in \mathcal{F}$ , let  $A_F$  be the event that the integers in  $F$  form an interval on this circle (ie, that they are all in a row next to each other). Since  $\mathcal{F}$  is intersecting, the number of sets  $F$  for which  $A_F$  can occur simultaneously is at most  $k$  (the maximum number of intervals of length  $k$  that can intersect is  $k$ ). Let



$1_F$  be the indicator that  $A_F$  occurs. Then for any outcome of the random process we have

$$\sum_{F \in \mathcal{F}} 1_F \leq k.$$

Therefore,

$$\mathbb{E} \left( \sum_{F \in \mathcal{F}} 1_F \right) = \sum_{F \in \mathcal{F}} \mathbb{P}(A_F) \leq k.$$

But for any  $F$ , the probability of  $A_F$  is exactly

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

Therefore

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

□



# 5

## Spectral graph theory

*The founders of Google computed the Perron-Frobenius eigenvector of the web graph and became billionaires.*

– Andries Brouwer and Willem Haemers

In this chapter we will learn spectral graph theory. Broadly speaking, the goal of spectral graph theory is to associate a matrix with a graph, and then use properties of the matrix to deduce properties of the graph. If you like linear algebra, you are in luck! Spectral graph theory is a lot of linear algebra. If you don't like linear algebra, you are also in luck! The matrices that we associate with a graph are all “very nice” and satisfy all of the properties that you would want a matrix to satisfy. This means that if you can remember that the matrices we are using satisfy certain nice properties, you can forget all of the other messy linear algebra that one needs for general matrices.

## 5.1 Linear Algebra Preliminaries

In this chapter, we will use properties of “nice” matrices frequently. Everything that you need to know is proved in a course like 241 or 242, and so I will not prove theorems about matrices here. Rather, we will take some basic linear algebra facts as a blackbox and use them frequently (the proofs are not hard, but it takes some set up). First, we will only be interested in matrices with entries that are real numbers.  $\mathbb{R}^{n \times m}$  will denote the set of such matrices that have  $n$  rows and  $m$  columns. Given a matrix  $A$  we will write  $A_{ij}$  to denote the entry in the  $i$ 'th row and  $j$ 'th column.

A *vector* is a matrix with only 1 column. Given a set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ , for any  $c_1, \dots, c_k \in \mathbb{R}$ , we say that

$$c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_k \mathbf{v}_k$$

is a *linear combination* of the vectors. We say that the *Span* of a set of vectors is the set of all linear combinations of those vectors

$$\text{Span}(\{\mathbf{v}_1, \dots, \mathbf{v}_k\}) = \left\{ \sum c_i \mathbf{v}_i : c_i \in \mathbb{R} \right\}.$$

We call a set of vectors  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  *linearly independent* if

$$\sum c_i \mathbf{v}_i = \mathbf{0}$$

implies that  $c_i = 0$  for all  $i$ . That is, none of the vectors can be written as a linear combination of the other ones. Given a subspace  $S$  of  $\mathbb{R}^n$ , a *basis* for  $S$  is a set of linearly independent vectors whose Span is all of  $S$ .

**Proposition 5.1.1.** *If  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  is a basis for  $S$ , then every vector in  $S$  can be written as a unique linear combination of vectors in  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ .*

*Proof.* If  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  is a basis for  $S$ , then any vector in  $S$  can be written as a linear combination of vectors in  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  by the definition of Span. Therefore, the only

thing to show is that this representation is unique. Let  $\mathbf{u} \in S$  and assume

$$\sum c_i \mathbf{v}_i = \sum b_i \mathbf{v}_i = \mathbf{u}.$$

Then we have

$$\sum (c_i - b_i) \mathbf{v}_i = \mathbf{0}.$$

Since  $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$  are linearly independent, this implies that  $(c_i - b_i) = 0$  for all  $i$ , ie the two representations of  $\mathbf{u}$  are the same.  $\square$

All bases for a set  $S$  have the same size (this is something you learn in 241), and the number of vectors in a basis for  $S$  is called the *dimension* of  $S$ .

A matrix in  $\mathbb{R}^{n \times n}$  is a function from  $\mathbb{R}^n$  to  $\mathbb{R}^n$ . ie, it takes a vector in  $\mathbb{R}^n$  and spits out another vector in  $\mathbb{R}^n$ . Furthermore, it is a consequence of matrix multiplication that this function is *linear*. That is, for vectors  $\mathbf{v}$  and  $\mathbf{u}$  we have

$$A(\mathbf{v} + \mathbf{u}) = A\mathbf{v} + A\mathbf{u},$$

and for any scalar  $c \in \mathbb{R}$  we have

$$A(c\mathbf{u}) = c(A\mathbf{u}).$$

Since  $A$  is a function, we may ask about its range. The co-domain is  $\mathbb{R}^n$ , but perhaps not all of the vectors are actually outputs of the function. Let  $S$  be the range of the function, that is

$$S = \{A\mathbf{v} : \mathbf{v} \in \mathbb{R}^n\}.$$

We define the *rank* of a matrix  $A$  to be the dimension of  $S$ . You learn in 241 that this is the same as the dimension of the Span of the rows of  $A$  (or of the columns).

For two vectors

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

we define their *dot product* to be

$$\mathbf{u} \cdot \mathbf{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n = \mathbf{u}^T \mathbf{v}.$$

Two vectors are said to be *orthogonal* if their dot product is 0 (in  $\mathbb{R}^n$ , this means that the two vectors are perpendicular to each other). Matrix multiplication can be defined via the dot product: for  $A, B \in \mathbb{R}^{n \times n}$ , we have

$$(AB)_{ij} = (i\text{'th row of } A) \cdot (j\text{'th column of } B).$$

Now we define something we will use constantly in this chapter, eigenvalues and eigenvectors. Given a matrix  $A$ , a nonzero vector  $\mathbf{x}$  is said to an *eigenvector* for  $A$  if there is a real number  $\lambda$  such that

$$A\mathbf{x} = \lambda\mathbf{x}.$$

$\lambda$  is called an *eigenvalue* of  $A$ . Another important fact (learned in 241) is that if a matrix is symmetric, then it has a set of  $n$  linearly independent eigenvectors (in fact, we can assume that the vectors are pairwise orthogonal). We say that the *trace* of a matrix  $A$  is the sum of its diagonal entries. If  $A$  is symmetric, then there is a set of eigenvectors and eigenvalues such that

$$A\mathbf{x}_i = \lambda_i\mathbf{x}_i$$

for  $1 \leq i \leq n$ . In this case (another fact from 241), we have  $\text{trace}(A) = \sum_{i=1}^n \lambda_i$ .

## 5.2 The adjacency matrix

*Can one hear the shape of a drum?*

– Mark Kac

*Can you hear the shape of a graph?*

– Fan Chung

We now study the most common matrix associated with a graph, the *adjacency matrix*. Given a graph  $G$  on  $n$  vertices we define a matrix  $A = A(G)$  which is an  $n$  by  $n$  matrix. Its rows and columns are indexed by the vertex set (in the same order for the rows as the columns), and the entries of  $A$  are given by

$$A_{ij} = \begin{cases} 1 & v_i \text{ is adjacent to } v_j \\ 0 & v_i \text{ is not adjacent to } v_j \end{cases}$$

Note that  $A$  is a symmetric matrix, since  $v_i$  is adjacent to  $v_j$  if and only if  $v_j$  is adjacent to  $v_i$ . Therefore, there is an orthogonal set of eigenvectors  $\mathbf{x}_1, \dots, \mathbf{x}_n$  corresponding to eigenvalues  $\lambda_1, \dots, \lambda_n$ . We will always order the indices so that

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

Given a graph  $G$ , we call the multiset of its eigenvalues the *spectrum* of  $G$

$$\text{Spec}(G) = \{\lambda_1, \dots, \lambda_n\}.$$

Further, if  $G$  has no loops, then the diagonal entries of  $A$  are all 0, and so  $\text{trace}(A) = \sum \lambda_i = 0$ . Now we come to the central question in spectral graph theory:

**Question 1.** *Given the spectrum of a graph  $G$ , what properties of  $G$  can we determine?*

In particular, given the spectrum of a graph, can you determine what the graph is? This is what is meant by “hearing the shape of a graph”. One graph property that we can immediately determine from the spectrum is the number of vertices in the graph. eg, if I tell you that I have a graph and its spectrum is

$$\{-1, -1, -1, -1, 4\}$$

then you can immediately tell me that my graph has 5 vertices. What else can we say about it? First we need to investigate the adjacency matrix a bit more:

**Theorem 5.2.1.** *The matrix  $A^k$  counts walks of length  $k$  in the graph. Specifically  $(A^k)_{ij}$  is the number of walks of length  $k$  from vertex  $v_i$  to vertex  $v_j$ .*

*Proof.* The proof is by induction on  $k$ . The number of walks of length 1 from  $v_i$  to  $v_j$  is 1 if  $i$  and  $j$  are adjacent and is 0 otherwise. By the definition of  $A$ , the base case is proved. Now assume it is true for  $A^k$ . Then

$$(A^{k+1})_{ij} = (A^k A)_{ij} = (\text{the } i\text{'th row of } A^k) \cdot (\text{the } j\text{'th column of } A) = \sum_{w=1}^n (A^k)_{iw} A_{wj}.$$

Now if  $v_w$  is not adjacent to  $v_j$  then the summand is 0. Otherwise by induction, this summand is equal to the number of walks of length  $k$  from  $v_i$  to  $v_w$ . Since in this case  $v_w \sim v_j$ , for each  $k$  walk from  $v_i$  to  $v_w$ , there is a  $k + 1$  walk from  $v_i$  to  $v_j$ . Since we sum over the whole vertex set, this counts all of the  $k$ -walks.  $\square$

We claim that this theorem tells us other information about the graph given its spectrum. Note that if  $\lambda$  is an eigenvalue of  $A$ , then  $\lambda^k$  is an eigenvalue of  $A^k$ . Therefore, if the spectrum of  $A$  is  $\{\lambda_1, \dots, \lambda_n\}$ , then the spectrum of  $A^2$  is  $\{\lambda_1^2, \dots, \lambda_n^2\}$ . But the matrix  $A^2$  counts walks of length 2 in the graph, and so for any  $i$  we have  $(A^2)_{ii} = \text{number of walks of length 2 from } v_i \text{ to itself} = d(v_i)$ . And so

$$\text{trace}(A^2) = \sum_{i=1}^n d(v_i) = 2e(G).$$



Therefore, knowing the spectrum of  $G$  determines the number of edges in  $G$ . So given a graph with spectrum  $\{-1, -1, -1, -1, 4\}$  we know that it has 5 vertices and

$$\frac{1}{2}((-1)^2 + (-1)^2 + (-1)^2 + (-1)^2 + 4^2) = 10$$

edges. What graph is it?

It is often useful to use the following characterization of  $\lambda_1$ :

$$\lambda_1 = \max_{x^T x \neq 0} \frac{x^T A x}{x^T x}.$$

Using this we can show that the spectral radius of a graph is between its average degree and its maximum degree.

**Theorem 5.2.2.** *Let  $G$  be a graph with maximum degree  $\Delta$  and average degree  $\bar{d} = \frac{2e(G)}{n}$ . Then*

$$\bar{d} \leq \lambda_1 \leq \Delta.$$

*Proof.* Since

$$\lambda_1 = \max_{x^T x \neq 0} \frac{x^T A x}{x^T x}.$$

we have

$$\lambda_1 \geq \frac{\mathbf{1}^T A \mathbf{1}}{\mathbf{1}^T \mathbf{1}},$$

where  $\mathbf{1}$  represents the all ones vector. But

$$\frac{\mathbf{1}^T A \mathbf{1}}{\mathbf{1}^T \mathbf{1}} = \frac{2e(G)}{n},$$

so the lower bound is proved. To see the upper bound, let  $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$ . Let

$$\mathbf{x}_1 = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

and assume that  $x_z$  is the largest entry (if there is more than one largest entry, choose of those arbitrarily). Then  $A\mathbf{x}_1 = \lambda_1\mathbf{x}_1$  is a system of equations, and looking at the one that corresponds to vertex  $z$  we have

$$\lambda_1 x_z = \sum_{v \sim z} x_v \leq \sum_{v \sim z} x_z = x_z d(z) \leq x_u \Delta.$$

And so  $\lambda_1 \leq \Delta$ . □

### 5.3 Short proofs of old results using spectral graph theory

In this section we use the previous set up to give short proofs of several old results. Using the inequality  $\lambda_1 \geq \frac{2e(G)}{n}$  will recover and sometimes strengthen these results. We use the same set up as before:  $G$  will be a graph and  $A$  will be its adjacency matrix.  $A$  will have eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  and  $\mathbf{x}$  will be an eigenvector corresponding to  $\lambda_1$ . We will normalize  $\mathbf{x}$  so that it has infinity norm 1 (ie its maximum entry is 1), and we will let  $z$  be a vertex chosen so that  $\mathbf{x}_z = 1$ . The eigenvector eigenvalue equation gives that for any vertex  $u$ ,

$$\lambda_1 \mathbf{x}_u = \sum_{v \sim u} \mathbf{x}_v.$$

Multiplying both sides of this equation by  $\lambda_1$  and applying the above equation again gives

$$\lambda_1^2 \mathbf{x}_u = \sum_{v \sim u} \lambda_1 \mathbf{x}_v = \sum_{v \sim u} \sum_{w \sim v} \mathbf{x}_w.$$

In particular, applying the above equation for  $u = z$  and noting that  $\mathbf{x}_w \leq 1$  gives

$$\lambda_1^2 = \sum_{v \sim z} \sum_{w \sim v} \mathbf{x}_w = \sum_{v \sim z} \sum_{\substack{w \sim v \\ w \in N(z)}} \mathbf{x}_w + \sum_{v \sim z} \sum_{\substack{w \sim v \\ w \notin N(z)}} \mathbf{x}_w \leq 2e(N(z)) + e(N(z), V(G) \setminus N(z)) \tag{5.1}$$

where the last inequality follows because each eigenvector entry is at most 1, and because each eigenvector entry appears at the end of a walk of length 2 from  $z$ : each edge with both endpoints in  $N(z)$  is the second edge of a walk of length 2 from  $z$  exactly twice and each edge with only one endpoint in  $N(z)$  is the second edge of a walk of length 2 from  $z$  exactly once. Using this setup we give short proofs of the following known results:

- Mantel's Theorem.
- Stanley's Bound (1987): if  $G$  is a graph with  $m$  edges and  $\lambda_1$  is the spectral radius of its adjacency matrix, then  $\lambda_1 \leq \frac{1}{2}(-1 + \sqrt{1 + 8m})$ .
- A long-standing conjecture of Erdős (1975) is that every triangle-free graph may be made bipartite with the removal of at most  $n^2/25$  edges. We show that the conjecture is true for graphs with at least  $n^2/5$  edges, first proved by Erdős, Faudree, Pach, and Spencer (1988).
- If  $G$  is a  $K_{2,t}$ -free graph and the spectral radius of its adjacency matrix is  $\lambda_1$ , then  $\lambda_1 \leq 1/2 + \sqrt{(t-1)(n-1) + 1/4}$ . This was originally proved by Nikiforov (2010) and is a spectral strengthening of the Kővari-Sós-Turán Theorem applied to  $\text{ex}(n, K_{2,t})$ .
- An improvement of the Stanley Bound (Hong, 1998) and some variants of it when one forbids cycles of length 3 and 4 (Nikiforov 2007 and Nosal 1970).
- An upper bound on the spectral radius of a graph based on local structure, first proved by Favaron, Mahéo, and Saclé (1993).

**Theorem 5.3.1** (Mantel's Theorem). *Let  $G$  be a triangle-free on  $n$  vertices. Then  $G$  contains at most  $\lfloor n^2/4 \rfloor$  edges. Equality occurs if and only if  $G = K_{\lfloor n/2 \rfloor \lfloor n/2 \rfloor}$ .*

*Proof.* If  $G$  is triangle-free, then  $e(N(z)) = 0$ . Using  $\lambda_1 \geq \frac{2e(G)}{n}$  and (5.1) gives

$$\frac{4(e(G))^2}{n^2} \leq e(N(z), V(G) \setminus N(z)) \leq \left\lceil \frac{n}{2} \right\rceil \left\lfloor \frac{n}{2} \right\rfloor.$$

Equality may occur only if  $e(N(z), V(G) \setminus N(z)) = \lfloor n^2/4 \rfloor$ . The only bipartite graph with this many edges is  $K_{\lfloor n/2 \rfloor \lceil n/2 \rceil}$ , and thus  $K_{\lfloor n/2 \rfloor \lceil n/2 \rceil}$  is a subgraph of  $G$ . But  $G$  is triangle-free, and so  $G = K_{\lfloor n/2 \rfloor \lceil n/2 \rceil}$ .

□

**Theorem 5.3.2** (Stanley's Bound). *Let  $G$  have  $m$  edges. Then*

$$\lambda_1 \leq \frac{1}{2} \left( -1 + \sqrt{1 + 8m} \right).$$

*Equality occurs if and only if  $G$  is a clique and isolated vertices.*

*Proof.* Using (5.1) gives

$$\lambda_1^2 = \sum_{z \sim v} \sum_{\substack{v \sim w \\ w \neq z}} \mathbf{x}_w + \sum_{v \sim z} 1 \leq 2(m - d_z) + d_z \leq 2m - \lambda_1,$$

where the last inequality is because  $\lambda_1 \leq d_z$ . The result follows by the quadratic formula. Examining (5.1) shows that equality holds if and only if  $E(G)$  is contained in the closed neighborhood of  $z$ ,  $d_z = \lambda_1$ , and for each  $w \sim z$ ,  $\mathbf{x}_w = 1$ . Since  $z$  was chosen arbitrarily amongst vertices of eigenvector entry 1, any vertex of eigenvector entry 1 must contain  $E(G)$  in its closed neighborhood. Thus  $G$  is a clique plus isolated vertices.

□

**Theorem 5.3.3** (Erdős-Faudree-Pach-Spencer). *Let  $G$  be a triangle-free graph on  $n$  vertices with at least  $n^2/5$  edges. Then  $G$  can be made bipartite by removing at most  $n^2/25$  edges.*

*Proof.* Let  $G$  be triangle-free with  $m$  edges, and let  $\text{MaxCut}(G)$  denote the size of a maximum cut in  $G$ . So we are trying to show that  $m - \text{MaxCut}(G) \leq n^2/25$ . Since  $G$  is triangle-free,  $N(z)$  induces no edges. Thus by (5.1)

$$\frac{4m^2}{n^2} \leq \lambda_1^2 \leq e(N(z), V(G) \setminus N(z)) \leq \text{MaxCut}(G).$$

Let  $g(m) = m - \frac{4m^2}{n^2}$ . The function  $g(m)$  is decreasing for  $m \geq \frac{n^2}{8}$ , and  $g(n^2/5) = n^2/25$ , which implies the result.  $\square$

**Theorem 5.3.4** (Nikiforov). *Let  $G$  be a  $K_{2,t}$ -free graph of order  $n$  and spectral radius  $\lambda_1$ . Then*

$$\lambda_1 \leq 1/2 + \sqrt{(t-1)(n-1) + 1/4}.$$

*Noting that  $\lambda_1 \geq \frac{2e(G)}{n}$  implies the Kővari-Sós-Turán Theorem applied to  $\text{ex}(n, K_{2,t})$ .*

*Proof.* Let  $w$  be a vertex not equal to  $z$ . Since  $G$  is  $K_{2,t}$ -free, there are at most  $t-1$  walks of length 2 from  $z$  to  $w$ . Therefore, by (5.1)

$$\begin{aligned} \lambda_1^2 &= d_z + \sum_{v \sim z} \sum_{\substack{w \sim v \\ w \in N(z)}} \mathbf{x}_w + \sum_{v \sim z} \sum_{\substack{w \sim v \\ w \notin N(z) \\ w \neq z}} \mathbf{x}_w \\ &\leq d_z + (t-1) \sum_{w \in N(z)} \mathbf{x}_w + (t-1) \sum_{\substack{w \notin N(z) \\ w \neq z}} \mathbf{x}_w \\ &= d_z + \sum_{w \in N(z)} \mathbf{x}_w + (t-2) \sum_{w \in N(z)} \mathbf{x}_w + (t-1) \sum_{\substack{w \notin N(z) \\ w \neq z}} \mathbf{x}_w \\ &= d_z + \lambda_1 + (t-2) \sum_{w \in N(z)} \mathbf{x}_w + (t-1) \sum_{\substack{w \notin N(z) \\ w \neq z}} \mathbf{x}_w \\ &\leq d_z + \lambda_1 + (t-2)d_z + (t-1)(n - d_z - 1). \end{aligned}$$

Applying the quadratic formula yields the result.  $\square$

The next three theorems are variants of Stanley's edge bound.

**Theorem 5.3.5** (Nosal (1970)). *If  $G$  is triangle free with  $m$  edges and spectral radius  $\lambda_1$ , then  $\lambda_1 \leq \sqrt{m}$ .*

*Proof.* If  $G$  is triangle-free, then  $e(N(z)) = 0$ . (5.1) implies

$$\lambda_1^2 \leq e(N(z), V(G) \setminus N(z)) \leq m.$$

□

**Theorem 5.3.6** (Nikiforov (2007)). *Let  $G$  be an  $n$ -vertex graph of girth at least 5 and spectral radius  $\lambda_1$ . Then  $\lambda_1 \leq \sqrt{n-1}$ .*

*Proof.* Since  $G$  is triangle and quadrilateral-free,  $e(N(z)) = 0$  and for any  $w \in V(G) \setminus \{z \cup N(z)\}$   $|N(w) \cap N(z)| \leq 1$ . Therefore  $e(N(z), V(G) \setminus N(z)) \leq d_z + (n - d_z - 1)$ . (5.1) gives  $\lambda_1^2 \leq n - 1$ . □

Using  $\lambda_1 \leq \Delta(G)$ , we have for  $G$  of girth at least 5,  $\lambda_1 \leq \min\{\Delta, \sqrt{n-1}\}$ . Nikiforov (2007) characterizes the cases of equality. We leave the characterization of equality using our proof to the reader.

**Theorem 5.3.7** (Hong). *Let  $G$  be a connected graph on  $m$  edges with spectral radius  $\lambda_1$ , then*

$$\lambda_1 \leq \sqrt{2m - n + 1}.$$

*Equality occurs if and only if  $G$  is either a complete graph or a star.*

*Proof.* Since  $G$  is connected, every vertex in  $V(G) \setminus \{z \cup N(z)\}$  has degree at least 1. Therefore, at least  $n - d_z - 1$  edges contribute at most 1 to the sum in (5.1). This gives

$$\lambda_1^2 \leq d_z + 2e(N(z)) - (n - d_z - 1) \leq 2m - n + 1.$$

Equality occurs if and only if for all  $u \in V(G) \setminus \{z \cup N(z)\}$ ,  $d_u = 1$  and for any walk of length 2 starting at  $z$  and ending at  $u$ ,  $\mathbf{x}_u = 1$ . These conditions together imply

that  $V(G) = z \cup N(z)$ . Now, if there are any edges in  $N(z)$ , then both endpoints must have eigenvector entry 1. If  $\mathbf{x}_w = 1$  and  $N(w) \subset \{z \cup N(z)\}$ , then  $N(w)$  must equal  $V(G) \setminus \{w\}$ . Therefore  $G$  is either a clique or a star.  $\square$

Finally, we note that some of the above theorems are corollaries of the following bound by Favaron, Mahéo, and Saclé. We prove (a stronger version of) their theorem immediately from (5.1)

**Theorem 5.3.8** (Favaron-Mahéo-Saclé). *Let  $G$  be a graph with spectral radius  $\lambda_1$ . For  $i \in V(G)$  let  $s_i$  be the sum of the degrees of the vertices adjacent to  $i$ . Then*

$$\lambda_1 \leq \max_i \sqrt{s_i}.$$

*Proof.* Since  $2e(N(z)) + e(N(z), V(G) \setminus N(z)) = s_z$ , we have immediately from (5.1)

$$\lambda_1^2 \leq s_z \leq \max_{i \in V(G)} s_i.$$

$\square$

## 5.4 The Graham-Pollak Theorem

*Beautiful graphs are rare. And so are gems like this proof.*

– Babai and Frankl

Originally motivated by a problem of loop switching in networks, Graham and Pollak became interested in partitioning the edge set of a multigraph by complete bipartite subgraphs (henceforth *bicliques*). If  $G$  is a finite, loopless multigraph, the *biclique partition number*, denoted  $\text{bp}(G)$ , is the minimum number of bicliques whose edge sets partition  $E(G)$ . Since every edge is a biclique, this parameter is well-defined and finite. Graham and Pollak showed that a problem on loop switching is equivalent to partitioning a multigraph, and in the process proved their celebrated theorem.

**Theorem 5.4.1** (Graham-Pollak Theorem). *The edge set of the complete graph on  $n$  vertices cannot be partitioned into fewer than  $n - 1$  complete bipartite subgraphs.*

As the edges of  $K_n$  can be partitioned into  $n - 1$  bicliques using edge-disjoint stars (there are also many other ways, exponentially many in fact), the Graham-Pollak Theorem gives the result  $\text{bp}(K_n) = n - 1$ . Since Graham and Pollak's result, many other proofs of this fact have been discovered. Though the result is purely combinatorial, most of the proofs are algebraic, including proofs by G.W. Peck, Tverberg, and Vishwanathan. Vishwanathan also discovered a proof that replaces linear algebraic techniques by using the pigeon-hole principle in a way that does not necessitate the use of an underlying field. However, his proof mimics Gaussian elimination and uses intermediate structures of large size (on the order of  $n^n$ ). He asked whether there was a "better" combinatorial proof.

We now describe a beautiful result, attributed Witsenhausen [?], which gives the Graham-Pollak Theorem as a corollary.

**Theorem 5.4.2** (Witsenhausen, 1980s). *Let  $G$  be a finite, loopless graph, and  $A$  its adjacency matrix. Then, if  $n_+(A)$  and  $n_-(A)$  denote the number of positive eigenvalues and negative eigenvalues of  $A$  respectively,*

$$\text{bp}(G) \geq \max(n_+(A), n_-(A)).$$

Since  $K_n$  has eigenvalue  $-1$  with multiplicity  $n - 1$ , the Graham-Pollak Theorem is a corollary.

*Proof.* Assume the edge set of a graph  $G$  is partitioned into  $\text{bp}(G)$  bicliques. If  $S$  is a subset of the vertices of  $G$ , then the characteristic vector of  $S$  is the  $n$ -dimensional  $(0, 1)$  column vector whose  $i$ -th position equals 1 if vertex  $i$  is in  $S$  and equals 0 otherwise. Denote by  $u_i$  and  $v_i$  the characteristic vectors of the partite sets of the



$i$ -th biclique of our decomposition. Define  $D_i = u_i v_i^T + v_i u_i^T$ . Then  $D_i$  is the adjacency matrix of the  $i$ -th biclique as a subgraph of  $G$ , and  $A = \sum_{i=1}^{\text{bp}(G)} D_i$ . Let

$$W = \text{Span}\{w \in \mathbb{R}^n \mid w^T u_i = 0, \forall 1 \leq i \leq \text{bp}(G)\}$$

$$P = \text{Span}\{\text{Eigenvectors of the positive eigenvalues of } A\}.$$

Since  $W$  is made up of  $n$ -dimensional vectors that are all orthogonal to  $\text{bp}(G)$  vectors, we have that  $\dim(W) \geq n - \text{bp}(G)$ . On the other hand, since  $p^T A p > 0$  for all nonzero  $p \in P$ , we have that  $W \cap P = \{0\}$ . Therefore

$$\dim(W) \leq n - \dim(P) = n - n_+(A).$$

It follows that  $n - \text{bp}(G) \leq \dim(W) \leq n - n_+(A)$  which implies that  $\text{bp}(G) \geq n_+(A)$ . The argument for  $n_-(A)$  follows similarly. Thus  $\text{bp}(G) \geq \max\{n_+(A), n_-(A)\}$ .  $\square$

The two most interesting open problems, in the opinion of the author, are the following.

**Open Problem 1.** *What is the minimum number of bicliques necessary to cover every edge of  $K_n$  at least once and at most twice? That is, what is  $\text{bp}_2(K_n)$ ?*

The best known bounds for this problem are given by  $\sqrt{n-1} \leq \text{bp}_2(K_n) \leq \lceil \sqrt{n} \rceil + \lfloor \sqrt{n} \rfloor - 2$  (cf Alon 1997).

**Open Problem 2.** *What is the minimum number of complete  $r$ -partite  $r$ -uniform hypergraphs necessary to partition the edge set of the complete  $r$ -uniform hypergraph?*

This question seems to be extremely difficult, and the best bounds, due to Cioabă, Kündgen, and Verstraëte in 2009 and Leader and Tan in 2017, are far apart. It would be nice to use an eigenvalue approach here as well, but the eigenvalue theory for higher order tensors is still not fully developed.

## 5.5 The Expander-Mixing Lemma

In this section we discuss how the second and last eigenvalue of a graph can tell us about the structure of the graph. We already saw how the first eigenvalue can tell us about the maximum and average degree. Let's say a graph is  $d$  regular: this property can be deduced from the spectrum (if and only if  $d = \lambda_1 = \frac{1}{n} \sum \lambda_i^2$ ). But there are many very different looking graphs which are all  $d$  regular, say two disjoint copies of  $K_{n/2}$  versus a  $K_{n/2, n/2}$ . Both of these are also very different from a "random regular graph with degree  $n/2$ " (whatever that means). Can we detect these structural differences from the spectrum? The answer is yes.

Given subsets  $S, T \subset V(G)$ , we define

$$e(S, T) = |\{(u, v) : u \in S, v \in T, uv \in E(G)\}|.$$

That is,  $e(S, T)$  denotes the number of edges between  $S$  and  $T$  where if  $S$  and  $T$  overlap and there is an edge with both endpoints in both of them, it is counted twice. If  $G$  is a  $d$ -regular graph and we were to choose  $S$  of size  $|S|$  and  $T$  of size  $|T|$  uniformly at random, then the probability for a vertex to be in  $S$  is  $\frac{|S|}{n}$  and for it to be in  $T$  is  $\frac{|T|}{n}$ , and therefore by linearity of expectation we have

$$\mathbb{E}[e(S, T)] = \sum_{uv \in E(G)} \frac{|S|}{n} \frac{|T|}{n} + \frac{|T|}{n} \frac{|S|}{n} = \frac{dn}{2} \cdot \frac{2|S||T|}{n^2} = \frac{d|S||T|}{n}.$$

The Expander-Mixing Lemma gives us a bound on how far away two sets in a graph can be from this average:

**Theorem 5.5.1** (Expander Mixing Lemma: Alon, Chung). *Let  $G$  be a  $d$ -regular graph and  $S, T \subset V(G)$ . Then*

$$\lambda_n \sqrt{|S||T| \left(1 - \frac{|S|}{n}\right) \left(1 - \frac{|T|}{n}\right)} \leq e(S, T) - \frac{d|S||T|}{n} \leq \sqrt{|S||T| \left(1 - \frac{|S|}{n}\right) \left(1 - \frac{|T|}{n}\right)}$$

*Proof.* We prove the lower bound here. The upper bound is the same proof. Let  $A$  be the adjacency matrix of  $G$  and let  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  be its eigenvalues. Assume that  $A\mathbf{x}_i = \lambda_i\mathbf{x}_i$  and that the eigenvectors are orthonormal, ie  $\mathbf{x}_i^T\mathbf{x}_j = 0$  if  $i \neq j$  and 1 if  $i = j$ .

Let  $S, T \subseteq V(G)$ , and let  $\mathbf{1}_S$  ( $\mathbf{1}_T$ ) denote the characteristic vector of  $S$  (of  $T$ ). Then  $\mathbf{1}_S, \mathbf{1}_T \in \mathbb{R}^n$  and we will now show that  $e(S, T) = \mathbf{1}_S^T A \mathbf{1}_T$ .

$$x^T Ay = x^T (Ay) = \sum_i x_i (Ay)_i = \sum_i x_i \sum_j A_{ij} y_j = \sum_{i,j} A_{ij} x_i y_j$$

In the above equation,  $A_{ij} \neq 0$  precisely when  $i \sim j$  (in this case  $A_{ij} = 1$ ), hence  $\mathbf{1}_S^T A \mathbf{1}_T$  counts precisely the number of edges between  $S$  and  $T$ .

We can write  $\mathbf{1}_S = \sum \alpha_i x_i$  where  $\alpha_i = \mathbf{1}_S^T x_i$ , and similarly  $\mathbf{1}_T = \sum \beta_i x_i$  where  $\beta_i = \mathbf{1}_T^T x_i$ . Note that  $\sum \alpha_i^2 = \mathbf{1}_S^T \mathbf{1}_S = |S|$  and  $\sum \beta_i^2 = \mathbf{1}_T^T \mathbf{1}_T = |T|$ . Since  $G$  is  $d$ -regular,  $x_1 = \frac{1}{\sqrt{n}} \mathbf{1}$ . Therefore  $\alpha_1 = \mathbf{1}_S^T x_1 = \frac{1}{\sqrt{n}} |S|$  and  $\beta_1 = \mathbf{1}_T^T x_1 = \frac{1}{\sqrt{n}} |T|$ . Then we have the following:

$$\begin{aligned} e(S, T) &= \mathbf{1}_S^T A \mathbf{1}_T = \left( \sum_i \alpha_i x_i \right)^T A \left( \sum_j \beta_j x_j \right) \\ &= \left( \sum_i \alpha_i x_i \right)^T \left( \sum_j \beta_j A x_j \right) \\ &= \sum_{i,j} (\alpha_i x_i) (\beta_j \lambda_j x_j) \\ &= \sum_i \lambda_i \alpha_i \beta_i \\ &= \lambda_1 \alpha_1 \beta_1 + \sum_{i=2}^n \lambda_i \alpha_i \beta_i \\ &= d \cdot \frac{1}{n} |S| |T| + \sum_{i=2}^n \lambda_i \alpha_i \beta_i. \end{aligned}$$

Hence we have

$$e(S, T) - d \frac{|S||T|}{n} = \sum_{i=2}^n \lambda_i \alpha_i \beta_i \geq \lambda_n \sum_{i=2}^n \alpha_i \beta_i \geq \lambda_n \sum_{i=2}^n |\alpha_i| |\beta_i|$$

where the last inequality follows because  $\lambda_n < 0$ . Now, using Cauchy-Schwarz gives

$$e(S, T) - d \frac{|S||T|}{n} \geq \lambda_n \left( \sum_{i=2}^n \alpha_i^2 \right)^{1/2} \left( \sum_{i=2}^n \beta_i^2 \right)^{1/2} = \lambda_n \left( |S| - \frac{|S|^2}{n} \right) \left( |T| - \frac{|T|^2}{n} \right).$$

Rearranging gives the result. □

As a corollary we have the very useful Hoffman Ratio bound:

**Theorem 5.5.2** (Hoffman ratio bound). *If  $G$  is a  $d$  regular graph then*

$$\alpha(G) \leq n \frac{-\lambda_n}{d - \lambda_n}.$$

*Proof.* If  $S$  is an independent set in  $G$  then  $e(S, S) = 0$ . Therefore, by the expander mixing lemma we have

$$-d \frac{|S|^2}{n} \geq \lambda_n |S| \left( 1 - \frac{|S|}{n} \right).$$

Rearranging gives  $|S| \leq n \frac{-\lambda_n}{d - \lambda_n}$ . □

## 5.6 The Hoffman-Singleton Theorem

In this section we ask the following optimization question: minimize the number of vertices in a graph subject to the constraints that it is regular of a fixed degree and has a fixed girth (length of the shortest cycle).

*Five is a magic number, so let us consider regular graphs of girth five.*

– Babai and Frankl

Let's say a graph  $G$  is  $d$ -regular and has girth 5. Since it has no cycles of length 3 or 4, then when we do a breadth first search from a vertex, all of the paths of length 2 from the root must lead to distinct vertices. Since the graph is  $d$ -regular, we have at least

$$1 + d + d(d - 1) = d^2 + 1$$

vertices in the graph. So we must have  $n \geq d^2 + 1$ . This idea generalizes to arbitrary girth:

**Theorem 5.6.1** (The Moore Bound). *Let  $G$  be a  $d$ -regular graph on  $n$  vertices with girth at least  $2k + 1$ . Then*

$$n \geq 1 + \sum_{i=0}^{k-1} d(d-1)^i.$$

Graphs attaining equality in this bound are called *Moore graphs*. Do Moore graphs exist? If  $d = 2$  then a cycle of length  $2k + 1$  is a Moore graph. If  $k = 1$ , then the complete graph is a Moore graph with girth 3 and degree  $n - 1$ . After this it gets harder: the Petersen graph is a 3 regular graph on  $3^2 + 1$  vertices with girth 5 and so it is a Moore graph. Hoffman and Singleton constructed a graph on  $7^2 + 1$  vertices which is regular of degree 7 and has girth 5. Their graph is similar in some sense to the Petersen graph, and so one may think that there is an infinite family of Moore graphs of girth 5. The following incredible theorem says that this is not true.

**Theorem 5.6.2** (Hoffman and Singleton, 1960). *If  $G$  is a  $d$ -regular graph with girth 5 on  $d^2 + 1$  vertices, then  $d \in \{2, 3, 7, 57\}$ .*

*Proof.* First we show that the graph is more regular than we might expect. Take a pair of vertices  $u$  and  $v$ . If  $u \sim v$  then their common neighborhood must be empty, otherwise  $G$  would have a triangle. What if  $uv \notin E(G)$ ? If we go back to the breadth first search argument, then since  $n = d^2 + 1$ , we must have that all of the vertices are

in a breadth first search two steps away from the root. Letting  $u$  be the root, since we assumed that  $u \not\sim v$ , then  $v$  must be in the bottom layer of the tree. This means that there is a unique path of length 2 from  $u$  to  $v$  and so their common neighborhood has size 1.

Now, the adjacency matrix has a 1 in position  $uv$  if and only if  $u \sim v$ . By the above discussion,  $A^2$  has a 1 in position  $uv$  if and only if  $u \not\sim v$ . Therefore we have

$$A^2 + A = (d - 1)I + J,$$

where  $J$  is the all ones matrix. Now, since  $G$  is regular, we have that the all ones vector is an eigenvector for it corresponding to eigenvalue  $d$ . Furthermore, we can assume that all of the other eigenvectors are orthogonal to this all ones vector. Let  $A\mathbf{x} = \lambda\mathbf{x}$  where  $\mathbf{x}$  is one of the other eigenvectors. Then we have

$$\lambda^2\mathbf{x} + \lambda\mathbf{x} - (d - 1)\mathbf{x} = (A^2 + A - (d - 1)I)\mathbf{x} = J\mathbf{x} = \mathbf{0}.$$

Multiplying both sides by  $\mathbf{x}$  gives that  $\lambda^2 + \lambda - (d - 1) = 0$  and therefore  $\lambda = \frac{-1 \pm \sqrt{4d - 3}}{2}$ . Note that this holds for any eigenvalue that does not have the all ones eigenvector, so this means that  $G$  has eigenvalue  $d$  with multiplicity 1 and the other two eigenvalues with some multiplicities  $m_1$  and  $m_2$ . Now the total number of eigenvalues is the number of vertices and so

$$1 + m_1 + m_2 = d^2 + 1.$$

Also, the trace of  $A$  is 0, so we must have

$$d + m_1 \left( \frac{-1 + \sqrt{4d - 3}}{2} \right) + m_2 \left( \frac{-1 - \sqrt{4d - 3}}{2} \right) = 0.$$

Rearranging gives

$$2d - (m_1 + m_2) + (m_1 - m_2)\sqrt{4d - 3} = 2d - d^2 + (m_1 - m_2)\sqrt{4d - 3} = 0.$$

Now,  $4d - 3$  is an integer, and so  $\sqrt{4d - 3}$  is either an integer or irrational. If  $\sqrt{4d - 3}$  is irrational, then we must have  $m_1 = m_2$  (since 0 is the right hand side of the equation and is rational). In this case, we have  $2d - d^2 = 0$  and so  $d = 2$ .

For the remainder of the proof, assume that  $\sqrt{4d - 3}$  is an integer, call it  $s$ . Then

$$d = \frac{s^2 + 3}{4}.$$

Substituting this expression for  $d$  gives

$$2 \left( \frac{s^2 + 3}{4} \right) - \left( \frac{s^2 + 3}{4} \right)^2 + (m_1 - m_2)s = 0,$$

or equivalently

$$8s^2 + 24 - (s^4 + 6s^2 + 9) + 16(m_1 - m_2)s = 0.$$

Rearranging gives

$$s(s^3 - 2s - 16(m_1 - m_2)) = 15.$$

Since  $s$  and  $m_1, m_2$  are integers, this means that  $s$  must divide 15, and so  $s \in \{1, 3, 5, 15\}$ . Therefore in this case  $d = (s^2 + 3)/4$  must satisfy  $d \in \{1, 3, 7, 27\}$ . If  $d = 1$  then the graph is a matching, and so we throw it away, finishing the proof of the theorem.  $\square$

To end these notes, we mention that it is an open problem to determine if the Moore graph of degree 57 exists!