Version Fall 2013

Combinatorics

Contents

1 Some introduction

Example 1.1. *A is a set of $n+1$ distinct numbers from $\{1, 2, ..., 2n\}$, then (a) there are two distinct numbers from A such that one divides the other; (b) there are two distinct numbers from A that are relatively prime.*

Example 1.2 (Erdös - Szekeres 1935)**.** *Given any sequence $a_1, a_2, ..., a_{mn+1}$ of $mn+1$ distinct numbers, one can always find a sub-sequence of length $m+1$ that is increasing, or a sub-sequence of length $n+1$ that is decreasing.*

Example 1.3. *Pick 5 red points from the vertices of a regular 12-gon, then 5 yellow points from the vertices of another regular 12-gon. Prove that there is a pair of isometric triangles, one formed by three red points, the other with three blue points.*

Example 1.4. *In the beginning there is a peg on the position $(0, 0)$. In each step one may remove a peg at (a, b) and put one peg on each of $(a+1, b)$ and $(a, b+1)$, provided both positions are empty. How can you play the game so that there will be no peg in the $x + y \leq 1$ area? How about $x + y \leq 2$, $x + y \leq 3$, etc.?*

Example 1.5. *Let E be a set of 8-shaped curves (two circles touch at onr point from outside) in the plane, such that no two curves share any point. Prove that $|E|$ is countable.*

Example 1.6 (Erdös - Rényi - Sós 1966)**.** *In a party any two person has exactly one common friend, then there are $2k+1$ persons in the party, one is everyone's friend, and others form k pair of friends.*

Conjecture 1.1 (Erdös - Gyárfás 1995)**.** *Every graph with minimum degree 3 has a cycle whose length is a power of 2.*

2 Basic counting

We first define our play ground.

Notation. *For any positive integer n,*

$$[n] := \{1, 2, ..., n\}.$$

We will always be happy to view the objects in different forms. For now, a subset of $[n]$ can be written by the usual set language, or by a 0-1 (or ± 1, or Y-B) vector of length n.

Definition 2.1. *A* lattice path *in the plane is one where we take one vertical or horizontal unit in each step, and always gets closer to the destination.*

How many lattice paths are there from $(0, 0)$ to $(4, 3)$? We can solve it with dynamic programming. Give a combinatorial explanation. How about from $(0, 0)$ to (r_1, r_2), from $(0, 0, 0)$ to (r_1, r_2, r_3)?

Notation. $\binom{[n]}{r}$ *as a set (of all r-element subsets of $[n]$) is defined as*

$$\binom{[n]}{r} = \{S \subseteq [n] : |S| = r\}.$$

And

$$\binom{n}{r} := \left|\binom{[n]}{r}\right|.$$

$\binom{n}{r}$, the number of ways to pick r persons from n persons, the number of ways to pick a r element set from $\{1, 2, \cdots, n\}$, or the number of 0-1 strings of length n with r 1's.
Calculating $\binom{n}{r}$. (1) directly; (2) recurrence and Pascal triangle (and lattice path).

Notation.

$$(n)_k = n(n-1)...(n-k+1).$$

We have

$$\binom{n}{r} = (n)_r/r! = \frac{n!}{r!(n-r)!}.$$

Clearly we have dozens of reasons that

Example 2.1.

$$\binom{n}{r} = \binom{n}{n-r}$$

For sets A and B, we define the set union $(A \cup B)$, intersection $(A \cap B)$, difference $(A - B)$, product $(A \times B)$, and powers (A^B) as usual. Each element of A^B is a mapping (function) from B to A, i.e.

$$A^B := \{f | f : B \to A\}$$

In this situation we may view B as the indices, and any function is choosing one element from A for each index. Hence the structure of $[n]^{[k]}$ is the same as $[n]^k$ – the k-tuples of numbers in $[n]$.

Exercise 2.1. *The number of mappings from $[k]$ to $[n]$ is n^k, among them $(n)_k$ are injections.*

We note that the binomial theorem says that

$$(1+x)^n = \sum_{i=0}^{n} \binom{n}{i} x^i.$$

We may view the l.h.s. as $(1+x)(1+x)...(1+x)$, and the r.h.s. as its expansion, $\binom{n}{i}$ is indeed the number of ways we can get x^i from the expansion. This is the basic version of generating functions, which are just a series of placeholders.

Example 2.2. *Understand what does $(1+x)(1+x^2)(1+x^4)(1+x^8)(1+x^{16})$ tell us.*

Fact 2.1. *If p is prime and $0 < i < p$, then $\binom{p}{i} = 0 \pmod p$.*

Example 2.3. *Prove that $2^p - 2 \equiv 0 \pmod p$ for prime number p.*

Theorem 2.2 (Lucas 1878)**.** *Let p be a prime, $m, n \in \mathbf{N}$, write the numbers m and n in base p: $m = (m_k m_{k-1} ... m_0)$ and $n = (n_k n_{k-1} ... n_0)$; then*

$$\binom{m}{n} \equiv \prod_{0}^{k} \binom{m_i}{n_i}.$$

Equivalently,

$$\binom{m}{n} \equiv \binom{m \mod p}{n \mod p} \binom{\lfloor m/p \rfloor}{\lfloor n/p \rfloor}.$$

Proof. Let $n = tp + r$. Let $m = sp + r'$. Write out

$$\begin{aligned}\binom{m}{n} &= \frac{m(m-1)\cdots(m-r+1)}{n(n-1)\cdots(n-r+1)}\frac{(m-r)(m-r-1)\cdots(m-n+1)}{tp(tp-1)\cdots 1} \\ &\equiv \frac{m(m-1)\cdots(m-r+1)}{r!}\frac{A^t p^t s(s-1)\cdots(s-t+1)}{A^t p^t t!} \\ &\equiv \binom{r'}{r}\binom{s}{t}.\end{aligned}$$

Where $A = \prod_{i=1}^{p-1} 1$. In the last step, the first term is 0 if $r' < r$, but in any case the equation holds. □

Proof. (one more) Consider the polynomial in x over the field $\mathbf{Z}_p$: $P(x) := (1+x)^m$. $\mathcal{C}_{x^n} P = \binom{m}{n}$ (in $\mathbf{Z}_p$).
Note that $(1+x)^p = 1 + x^p$. So

$$P(x) = \prod_{i=0}^{k}(1+x)^{m_i p^i} = \prod_{i=0}^{k}(1+x^{p^i})^{m_i} = \prod_{i=0}^{k}\sum_{j=0}^{m_i}\binom{m_i}{j}x^{j\cdot p^i} = \prod_{i=0}^{k}\sum_{j=0}^{p-1}\binom{m_i}{j}x^{j\cdot p^i}.$$

Note that in the last form, there is only one way to get x^n. □

Use Lucas' theorem, consider $\mathbf{Z}_2$, we solve the following.

Example 2.4. *Find the number of odd entries in the n-th row of the Pascal triangle.*

More symmetric notation. Let $n = r_1 + r_2 + \cdots + r_t$. The multinomial number.

$$\binom{n}{r_1, r_2, \cdots, r_t} = \binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-\ldots-r_{k-1}}{r_k} = \frac{n!}{r_1! r_2! \cdots r_k!}$$

Consider the expansion of $(x+y+z)^n$, for any $r_1+r_2+r_3 = n$, the cooefficient of the term $x^{r_1}y^{r_2}z^{r_3}$ is $\binom{n}{r_1,r_2,r_3}$.

Example 2.5. $\Sigma_{r=0}^{n}\binom{n}{r} = 2^n$.

Solution. Analytic: Consider the expansion of $(1+x)^n$ and use $x = 1$.
Combinatorial: Both sides are counting the number of subsets of $[n]$. □

Example 2.6. $\Sigma_{r=0}^{n}(-1)^r\binom{n}{r} = 0$.

Solution. Analytic: Use $x = -1$ in $(1+x)^n$.
Combinatorial: Define a mapping $f : 2^{[n]} \to 2^{[n]}$ as follows.

$$f(S) = \begin{cases} S - \{1\} & \text{if } 1 \in S \\ S + \{1\} & \text{if } 1 \notin S \end{cases}$$

Easy to check f is its own inverse, and it is a bijection between even-subsets and odd-subsets.
Equivalently, think any subset as a 0-1 vector, the mapping is simply to flip the first bit.
Another proof: As a student pointed out, use the Pascal triangle, both even subsets and odd subsets equals the sum of the previous row. This actually gives a more detailed picture for the combinatorial proof above. (Think about it.) □

Example 2.7. *Give a combinatorial proof for* $\Sigma_{r=0}^{n}\binom{n}{r}^2 = \binom{2n}{n}$.

Example 2.8. *Give a combinatorial proof for*

$$\sum_{k=0}^{n}\binom{k}{a}\binom{n-k}{b} = \binom{n+1}{a+b+1}.$$

Solution. L.h.s: partition all the choices by where is the $(a+1)$-st seleted person. □

Example 2.9. *Give a combinatorial proof for*

$$\sum_{k=0}^{n} k\binom{n}{k}^2 = n\binom{2n-1}{n-1}.$$

Solution. Pick n person from n girls and n boys, one of them is the captain who must be a girl. □

Example 2.10.

$$\Sigma_{r=0}^{n} r\binom{n}{r} = n2^{n-1}.$$

Algebraic: Take dirivative of $(1+x)^n$ at $x=1$, or direct algebraic manipulation.
Combinatorial: Both sides counting the sum of the size of all subsets.
This kind of counting in two ways is usually a nice trick.

Example 2.11. *Let $f(n)$ be the number of divisors of n, what is $f(n)$ and what is the average of the first N numbers $\frac{1}{N}\sum_{n\leq N} f(n)$?*

Example 2.12. *Number of ways to put n* identical *balls into r labeled boxes. Formally, the number of solutions to*

$$x_1 + x_2 + \cdots + x_r = n$$

, where $x_i \in \mathbf{N}_0$

Answer. $\binom{n+r-1}{r-1}$, combinatorial explanation.

Example 2.13. *The number of solutions to*

$$x_1 + x_2 + \cdots + x_r = n$$

, where $x_i \in \mathbf{N}$

Answer. $\binom{n-1}{r-1}$, combinatorial explanation.

Example 2.14. *We used $x=-1$ in the expansion of $(1+x)^n$ to get $\sum \binom{n}{2k}$, now use the roots of $z^3=1$ to compute*

$$\sum_{k=0}^{\lfloor n/3 \rfloor} \binom{n}{3k}.$$

Check that when $n=9$ the above equals $(2^9-2)/3$.

3 Catalan numbers

A Vašek style quotation: They are called Catalan numbers because they are not first discovered by Catalan.
Here is the problem from a letter from Euler to Goldbach in 1751.

Example 3.1. *In how many ways can one triangulate a labeled* $(n+2)$*-gon by* $n-1$ *diagonals?*

Let the answer be C_n, we have $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, $C_4 = 14$, and we will be glad to define $C_0 = 1$.

Exercise 3.1.

$$C_n = \sum_0^{n-1} C_i C_{n-1-i}$$

Definition 3.1. *The Catalan number* $C(n) := \frac{1}{n+1}\binom{2n}{n}$.

Theorem 3.1 (Euler 1751)**.** *The number of ways to triangulate an* $(n+2)$*-gon by* $n-1$ *diagonals is* $C(n)$*.*

Proof. Let $G(x)$ be the g.f. for $\{C_n\}$, expand G^2 we get $G^2(x) = (G(x)-1)/x$. So,

$$G(x) = \frac{1 \pm \sqrt{1-4x}}{2x}$$

One way or another we will pick the minus here.

$$2xG(x) = 1 - (1-4x)^{1/2}$$

and we get the answer by using Taylor. □

One may already know many solutions as well many problems that leads to the same answer. We briefly mention just a few.

Proof. (one more for the triangulation problem) We establish a mapping as follows.
For a triangulation of the $(n+2)$-gon $a_1a_2...a_{n+2}$, pick any of its edge or diagonal and give it a direction ($4n+2$ choices).
For a triangulation of the $(n+3)$-gon $a_1a_2...a_{n+3}$, pick any of its edge except a_1a_2 ($n+2$ choices).

For any animal of the first kind, we duplicate the picked edge and open the two copies from the tail (of the picked direction). For any one of the second kind, we collapse the picked edge e, assume e was in the triangle efg, then f and g now become merged into one. It is your job to mediate and see this is a one-one mapping between $C_n \times (4n+2)$ and $C_{n+1} \times (n+2)$. □

Example 3.2. *There are* $\binom{2n}{n}$ *lattie paths from* $(0,0)$ *to* (n,n)*. How many of them never go below the diagonal* $y = x$*?*

It is a little more convenient to turn our pictures 45 degrees (by viewing up and right $\to +1$ and -1). Define a walk from $(0,0)$ where each step takes either $(1,1)$ or $(1,-1)$. Restate the problem

Example 3.3. *There are* $\binom{2n}{n}$ *walks from* $(0,0)$ *to* $(2n,0)$*. How many of them never go below the* x*-axis?*

Solution. We count the number of "bad" walks, i.e., those go below the x axis at least once. Focus on the first time it does this, it reached the line $y = -1$, we take the walk from that point to the end, reflect it about the line $y = -1$. It becomes a walk ends at $(2n, -2)$.
It is (your time to mediate again) easy to see that this gives a bijection between the bad walks and all the walks ends at $(2n, -2)$. So the number of bad walks is $\binom{2n}{n-1}$. □

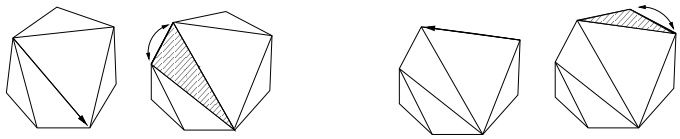

Figure 1: Some examples of the bijection.

4 Poset, lattice, and Möbius inversion

Definition 4.1. *A* partially ordered set (poset) *is a pair* $\mathcal{P} = (X, \preccurlyeq)$*, where* $\preccurlyeq$ *is a binary relation on* X *satisfying*
(1) reflexivity: $\forall x \in X$*,* $x \preccurlyeq x$*.*
(2) anti-symmetry: $\forall x, y \in X$*, if* $x \preccurlyeq y$ *and* $y \preccurlyeq x$*, then* $x = y$*.*
(3) transitivity: $\forall x, y, z \in X$*, if* $x \preccurlyeq y$ *and* $y \preccurlyeq z$*, then* $x \preccurlyeq z$*.*

We denote $x \prec y$ if $x \preccurlyeq y$ and $x \neq y$. Denote $[x, y]$ to be the *interval* $\{z : x \preccurlyeq z \preccurlyeq y\}$. We write $x \lessdot y$ if $[x, y] = \{x, y\}$ and we call y *covers* x.
Without specified otherwise, all the posets we consider here are finite.
We usually draw the *Hasse diagram* of a poset only showing its cover relations, draw a line between x and y (with y above x) if $x \lessdot y$.
It is clear from the definition of a poset that these are just the same animals as the reflexive, transitive closures of directed acyclic graphs (d.a.g.).
Some easy and some prominent examples of posets:

Example 4.1. *(1)* $(X, \preccurlyeq)$ *where* $X = \{a_1, a_2, ..., a_k, 0, 1\}$*, where* $0 \preccurlyeq x$ *and* $x \preccurlyeq 1$ *for any* $x \in X$*. i.e.*

$$\preccurlyeq = (\{0\} \times X) \cup (X \times \{1\}) \cup \{(x, x) : x \in X\}.$$

(2) $([n], \leqslant)$*, where* $\leqslant$ *is the usual integer order, is a chain.*
(3) $(2^{[n]}, \subseteq)$*, where* $\subseteq$ *is the usual set inclusion.*
(4) $([n], |)$*, where* $|$ *is the divisibility relation.*

It is illustrative to play with some small special cases of the example. In this section we suggest $k = 4$ for (1), $n = 5$ for (2), $n = 3$ for (3), and $n = 12$ for (4).

Definition 4.2. *Let* $\mathcal{P} = (X, \preccurlyeq)$ *be a poset, its* incidence algebra *is*

$$\mathcal{A}(\mathcal{P}) := \{\alpha | \alpha : X \times X \to \mathbf{R}, \alpha(x, y) = 0 \ \ \textit{whenever} \ \ x \not\preccurlyeq y\}.$$

i.e., $\mathcal{A}(\mathcal{P})$ *consists of matrices whose non-zero entries are all inside the* $\preccurlyeq$ *relation.*

Fact 4.1. *Let* $\mathcal{P}$ *be a poset,*
(1) If $\alpha, \beta \in \mathcal{A}(\mathcal{P})$*,* $c \in \mathbf{R}$*, then* $\alpha + \beta \in \mathcal{A}(\mathcal{P})$ *and* $c\alpha \in \mathcal{A}(\mathcal{P})$*.*
(2) If $\alpha, \beta \in \mathcal{A}(\mathcal{P})$*, then* $\alpha \cdot \beta \in \mathcal{A}(\mathcal{P})$*.*

Proof. (sketch) (1) follows directly from the definition of incidence algebra. (2) follows from the definition of the multiplication of matrices. □

We will often use the following notation.

Notation. *Let R be a subset of the universe U, define its* characteristic function *as*

$$\chi_R(x) = \begin{cases} 1 & \text{if } x \in R \\ 0 & \text{otherwise} \end{cases}$$

Note: Often U is in the form $(X_1 \times X_2 \times ... \times X_k)$, and R is a subset of U, i.e., a k-ary relation. For example, we have $\chi_{x \leq y}$, $\chi_{x|y}$, etc.
It is easy to check the following is an element in the incidence algebra

Definition 4.3. *Let $\mathcal{P} = (X, \preccurlyeq)$ be a poset, its ζ-function is defined as*

$$\zeta_P := \chi_{x \preccurlyeq y}.$$

Exercise 4.1. *Write ζ-function in a matrix form for the examples in Example 4.1. (Do this explicitly for the small numbers as we suggested.)*

For we people with the right background, it is easy to see ζ is has a inverse: Any poset (or a d.a.g.) has a linear extension (topological order). When we arrange the elements of X in this order, and write ζ in matrix form, then it is an upper-triangular matrix, with all 1's on the diagonal. So it has an inverse. Moreover, ζ has determinant 1, hence its inverse has all integer entries due to Cramer's formula. Thus we can define

Definition 4.4. *Let $\mathcal{P} = (X, \preccurlyeq)$ be a poset, its Möbius function is defined as*

$$\mu_P := \zeta_P^{-1}.$$

In fact we may solve μ explicitly. We want $\mu \cdot \zeta = I(:= \chi_{x=y})$. By the definition, it is equivalent to,

$$\forall x, y, \sum_{z \in X} \mu(x, z)\zeta(z, y) = \chi_{x=y}.$$

i.e.

$$\forall x, y, \sum_{z \preccurlyeq y} \mu(x, z) = \chi_{x=y}. \tag{1}$$

Example 4.1. $\mu(x,y) = 0$ *whenever* $x \not\preccurlyeq y$. *i.e.* $\mu \in \mathcal{A}(\mathcal{P})$.

Proof. Prove by contradiction, consider the violation pair (x,y) where y is minimal, and use (1). □

Hence, we may focus on $\mu(x,y)$ where $x \preccurlyeq y$, and (1) becomes

$$\forall x, y, \sum_{x \preccurlyeq z \preccurlyeq y} \mu(x,z) = \chi_{x=y}. \tag{2}$$

Similarly,

$$\forall x, y, \sum_{x \preccurlyeq z \preccurlyeq y} \mu(z,y) = \chi_{x=y}. \tag{3}$$

Let $\mathcal{P}$ be a poset and $\mu = \mu_P$. From (2) it is easy to see $\mu(x,x) = 1$; $\mu(x,y) = -1$ if $x \prec y$; $\mu(0,1) = k-1$ in case (1) of Example 4.1. And μ is inductively determined on $|[x,y]|$.

Exercise 4.2. *Write* μ*-function in a matrix form for the examples in Example 4.1. (Do this explicitly for the small numbers as we suggested. And suggest solutions for the general cases.)*

Fact 4.2 (Möbius inversion)**.** *Let* $\mathcal{P} = (X, \preccurlyeq)$ *be a finite poset,* $f, g : X \to \mathbf{R}$*, then*
(a).

$$\forall x, g(x) = \sum_{a \preccurlyeq x} f(a) \quad \Leftrightarrow \quad \forall x, f(x) = \sum_{a \preccurlyeq x} \mu(a,x) g(a).$$

(b).

$$\forall x, g(x) = \sum_{x \preccurlyeq a} f(a) \quad \Leftrightarrow \quad \forall x, f(x) = \sum_{x \preccurlyeq a} \mu(x,a) g(a).$$

Proof. Functions $X \to \mathbf{R}$ can be viewed as (column) vectors on $\mathbf{R}^{|X|}$. The inversion is nothing but $g^T = f^T\zeta \Leftrightarrow g^T\mu = f^T$, and $g = \zeta f \Leftrightarrow \mu g = f$. □

Definition 4.5. *Let* $\mathcal{P} = (X, \preccurlyeq_P)$ *and* $\mathcal{Q} = (Y, \preccurlyeq_Q)$ *be two posets, their* product *is*

$$\mathcal{P} \times \mathcal{Q} = (X \times Y, \preccurlyeq),$$

where $(x,y) \leq (x',y')$ *iff* $x \preccurlyeq_P x'$ *and* $y \preccurlyeq_Q y'$.

It is easy to check $\mathcal{P} \times \mathcal{Q}$ is again a poset, and

Fact 4.3.

$$\mu_{P\times Q}((x,y),(x',y')) = \mu_P(x,x')\mu_Q(y,y')$$

Proof. One may do induction on the size $[(x,y),(x',y')]$. Note that

$$|[(x,y),(x',y')]| = |[x,y]| \cdot |[x',y']|.$$

Or directly check if μ thus defined inverses ζ. Or, for people really like linear algebra, use tensor product. We provide the detail for the induction:
The basis, when $|[(x,y),(x',y')]|$ is 0 or 1, is trivial. Now suppose $x \neq x'$ or $y \neq y'$. So

$$\sum_{x\preccurlyeq x''\preccurlyeq x', y\preccurlyeq y''\preccurlyeq y'} \mu((x,y),(x'',y'')) = 0 = \sum_{x\preccurlyeq x''\preccurlyeq x'} \mu(x,x'') \cdot \sum_{y\preccurlyeq y''\preccurlyeq y'} \mu(y,y'')$$

Suppose $|[x,x']| = s$ and $|[y,y']| = t$, then $|[(x,y),(x',y')]| = st$. The l.h.s. has st terms. Expand the r.h.s., we also have st terms, and by induction, the term for each pair $(x'',y'') \neq (x',y')$ cancels. So we get $\mu_{P\times Q}((x,y),(x',y'))$ left on the l.h.s and $\mu_P(x,x')\mu_Q(y,y')$ on the right. □

It is clear the product and the above fact can be extended to $\mathcal{P}_1 \times \mathcal{P}_2 \times ... \times \mathcal{P}_k$. Immediately we explained the Möbius function on $(2^{[n]}, \subseteq)$. And the poset $(X, |)$, where X is all the divisors of $n = p_1^{t_1} p_2^{t_2} ... p_k^{t_k}$ is isomorphic to

$$([t_1+1], \leqslant) \times ... \times ([t_k+1], \leqslant),$$

so

$$\mu(1,n) = \begin{cases} (-1)^k & t_i = 1 \text{ for all } i \\ 0 & \text{if } t_i > 1 \text{ for some } i \end{cases}$$

Definition 4.6. *Let $\mathcal{P} = (X, \preccurlyeq)$ be a poset and $T \subseteq X$. $x \in T$ is called a* minimal element *if the only $y \in T$ and $y \preccurlyeq x$ is x itself. $x \in T$ is called the* minimum element *if $x \preccurlyeq y$ for all $y \in T$. Similarly we define* maximal element *and* maximum element.

Definition 4.7. *Let $\mathcal{P} = (X, \preccurlyeq)$ be a poset and $S \subseteq X$. $x \in X$ is an* upper bound *of S if $y \preccurlyeq x$ for all $y \in S$. An upper bound x_0 is called the* least upper bound *if it is minimum among the upper bounds of S. Similarly we define* lower bound *and* greatest lower bound.

Definition 4.8. *A* chain *in a poset is* $(x_1, x_2, ..., x_k)$ *such that* $x_i \prec x_{i+1}$ *for all* $i < k$. *An* antichain *in a poset is a set* S *of elements where no pairs are comparable, i.e.,* $\forall x, y \in S$, $x \not\prec y$. *An* ideal *(or a downwards closed set) is a set* I *such that for any* $x \in S$, $y \in I$ *whenever* $y \prec x$. *A* filter *(or a upwards closed set) is a set* F *such that for any* $x \in F$, $y \in I$ *whenever* $x \prec y$.

Note that there is a natural 1-1 correspondance between ideals and antichains, so is there between filters and antichains.

Definition 4.9. *A poset* $\mathcal{L} = (X, \preccurlyeq)$ *is a* lattice *if for any non-empty finite subset* $S \subseteq X$, *there is a greatest lower bound (*meet*)* $\bigwedge S$ *and a least upper bound (*join*)* $\bigvee S$.

Notation. *For* $S = \{x, y\}$, *we write* $x \wedge y$ *as their meet, and* $x \vee y$ *as their join.*

It is easy to check that if any two elements have meet and join, so does any finite non-empty subset. $\wedge$ and $\vee$ are commutative and associative. So we may write $\bigwedge\{x_1, x_2, ..., x_k\}$ as $x_1 \wedge x_2 ... \wedge x_k$ without confusion.
If $\mathcal{L} = (X, \preccurlyeq)$ is a finite lattice, then there is a smallest element $0_L := \bigwedge X$ and a largest element $1_L := \bigvee X$.

Theorem 4.4 (Weisner 1935)**.** *Let* $\mathcal{L}$ *be a finite lattice and* $a \neq 0_L$, *then*

$$\sum_{x: x \vee a = 1_L} \mu(0, x) = 0.$$

Proof. Using $a \neq 0_L$ and (2)

$$\sum_{y: a \preccurlyeq y} \mu(y, 1)(\sum_{x \preccurlyeq y} \mu(0, x)) = \sum_{y: a \preccurlyeq y} \mu(y, 1)0 = 0.$$

We can count it from x and use (3):

$$\begin{aligned}\sum_{x} \mu(0, x)(\sum_{y: a \preccurlyeq y, x \preccurlyeq y} \mu(y, 1)) &= \sum_{x} \mu(0, x)(\sum_{y: a \vee x \preccurlyeq y} \mu(y, 1)) \\ &= \sum_{x} \mu(0, x)\chi_{a \vee x = 1} \\ &= \sum_{a \vee x = 1} \mu(0, x).\end{aligned}$$

□

Use Weisner's theorem we have another way to compute the μ function for our examples.

Example 4.2. *In* $(2^{[n]}, \subseteq)$*, pick any singleton set, say,* $\{1\}$*, we get* $\mu(0, A) = (-1)^{|A|}$ *(by induction on* $|A|$*).*

Example 4.3. $([n], |)$ *is not a lattice, but* $(\{a : a|n\}, |)$ *is, pick any* $p|n$*, we get*

$$\mu(1, n) = \begin{cases} (-1)^k & \text{if } n \text{ is the product of } k \text{ distinct primes} \\ 0 & \text{if } n \text{ is divisible by the square of a prime} \end{cases}$$

by induction on n*.*

5 Principle of inclusion and exclusion

The PIE for 2 sets (events) and 3 sets and the following general PIE.

Theorem 5.1 (The principle of inclusion-exclusion)**.** *Let S be a set of size n, $E_i \subseteq S$ $(1 \leq i \leq r)$. For $M \subseteq [r]$, define*

$$E_M := \cap_{i \in M} E_i \quad \textit{and} \quad E(M) := |E_M|,$$

i.e., E_M is the intersection of the subsets with indices in M. (Note that $E_\emptyset = S$.) Then,

$$|S - \cup E_i| = \sum_{M \subseteq [r]} (-1)^{|M|} E(M).$$

If we further define

$$N_j := \sum_{|M|=j} E(M)$$

then

$$|S - \cup E_i| = \sum_{i=0}^{n} (-1)^i N_i = N - N_1 + N_2 - N3 + \cdots + (-1)^r N_r.$$

And the partial sum of the first terms of the r.h.s. alternates above and below the l.h.s.

Note that N_j is the sum of $\binom{r}{j}$ terms.

Proof. (by double counting) 2^r rows indicate all the possible M's, n columns for the elements of S, put a 1 on (M, x) if $x \in E_M$, then multiply the row for M by $(-1)^{|M|}$. Then focus on the contribution of each column. □

Proof. (by Möbius inversion) For each $M \subseteq [r]$, define

$$F_M := \bigcap_{i \in M} E_i \cap \bigcap_{i \notin M} \overline{E_i},$$

i.e., these elements whose membership indicator is exactly M. Let $F(M) := |F(M)|$. Then we have $\{F'_M : M \subseteq M'\}$ is a partition of E_M, and

$$\forall M, E(M) = \sum_{M \subseteq M'} F(M').$$

By Möbius inversion,

$$\forall M, F(M) = \sum_{M \subseteq M'} \mu(M, M') E(M').$$

In particular,

$$|S - \cup E_i| = F(\emptyset) = \sum_{M} (-1)^{|M|} E(M).$$

□

The above proofs can be straightforwardly generalized to probability spaces.

Theorem 5.2. *Let E_i $(1 \leq i \leq r)$ be events in a probability space. For $M \subseteq [r]$, define*

$$E_M := \cap_{i \in M} E_i$$

Then,

$$\mathsf{Pr}(\overline{\cup E_i}) = \sum_{M \subseteq [r]} (-1)^{|M|} \mathsf{Pr}(E_M).$$

Example 5.1 (de Montmort 1708 - 1713)**.** *(Derangements D_n) Number of permutations of $[n]$ where no one gets her own hat.*

Solution. Let S be the set of all $n!$ permutations. For each $1 \leq i \leq n$, let $E_i :=$ the set of permutations where i is mapped to itself.
For each $M \subseteq [n]$, E_M, defined as $\cap_{i \in M} E_i$, is exactly the set of permutations where all elements in M are fixed. So $E(M) = (n - |M|)!$. Therefore

$$\begin{aligned} D_n =& n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \binom{n}{3}(n-3)! + \cdots \\ =& n!(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!}) \\ \sim& n! e^{-1}. \end{aligned} \tag{4}$$

□

We can easily see that $D_n = nD_{n-1} + (-1)^n$, and

Exercise 5.1.

$$D_{n+1} = n(D_n + D_{n-1}).$$

Also find a combinatorial proof.

Example 5.2. *Find a combinatorial proof for*

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

Solution. Both side are counting the number of permutations of $[n]$, the left side partitions all the permutations by the exact number of fixed points. □

Solution. (second solution to the derangement problem) We use the *exponential generating function*

$$f(x) = \sum_{k} \frac{D_k}{k!} x^k$$

and we know

$$e^x = \sum_{k} \frac{x^k}{k!}.$$

So, expand the product $g(x) = e^x f(x)$, and use the previous example, the coefficient for x^n is

$$\mathcal{C}_n = \sum_{k=0}^{n} \frac{D_{n-k}}{k!(n-k)!} = \frac{1}{n!} \sum_{k} \binom{n}{k} D_{n-k} = 1.$$

So $e^x f(x) = (1-x)^{-1}$, and we get $f(x) = e^{-x}(1 + x + x^2 + ...)$. Expand we get the formula for D_n as in (4). □

Example 5.3. *Prove that*

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

Proof. Choose k girls from n boys and m girls. $E_i :=$ the i-th boy is chosen. □

A problem similar to the derangements. There are $2n$ seats around the table, n women are seated on every other position. Now comes their husbands (exactly one for each woman, luckily), count the number of ways none of the husbands sit next to his wife. Formally,

Example 5.4. *Number of permutation of* $[n]$ *such that* i *is not mapped to* i *nor* $(i-1)$ *(mod* n*).*

Solution. Instead of setting n E_i's, we will find it is nicer to set $2n$ of them. Let E_i be the set of permutations where i is mapped to $i-1$, and E_i' be the set where i is mapped to i. We put E_0, E_0', E_1, E_1', ..., E_n, E_n' on a circle in this order. It is easy to see that the intersection of any two adjacent sets on the circle is empty. (i cannot be sitting on position i and $i-1$ at the same time, and if i sits on position i, then $i+1$ cannot.) On the other hand, whenever k of the E_i's happen, as long as no two are adjacent on the circle, the position of k persons are specified, and all the others are free to sit anywhere.
Formally, call M good if there are no adjacent indices in M on the $2n$-circle. We have

$$E(M) = \begin{cases} (n-|M|)! & \text{if M is good} \\ 0 & \text{if M is bad} \end{cases}$$

So the problem is reduced to counting the number of good M's of size k, i.e., the number of ways to pick k points on a $2n$-circle where no two are adjacent, which is (exercise) $\binom{2n-k-1}{k-1} + \binom{2n-k}{k} = \frac{2n}{2n-k}\binom{2n-k}{k}$. So, our answer is

$$\sum_{k=0}^{n} (-1)^k (n-k)! \frac{2n}{2n-k} \binom{2n-k}{k}$$

. □

Two more combinatorial proofs.

Exercise 5.2. *Colour the integers* 1 *to* $2n$ *red or blue s.t. if* i *is red, then* $i-1$ *must be red. A simpler way to view this is to colour the left most* 0 *or more integers red, and the rest blue. So*

$$\sum_{k=0}^{n} (-1)^k \binom{2n-k}{k} 2^{2n-2k} = 2n+1.$$

Exercise 5.3. *Prove that for any* $0 \le k \le n$*,*

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

Hint: Similar to the derangement problem. But now we count the number of permutations where it is OK for any of the first k person to get her own hat.

Definition 5.1. *Let A be a $n \times n$ matrix, its* permanant *is defined as*

$$perm(A) := \sum_{\substack{n \text{ numbers} \\ \text{one from each row} \\ \text{and one from each column}}} \text{the product of these numbers}$$

i.e.

$$perm(A) := \sum_{\sigma \in S_n} \prod_i A_{i,\sigma(i)},$$

where S_n is the set of permutations on $[n]$.

Note that when A is a 0-1 matrix, permanant is the count of permutations that having all 1's. Such a matrix can be viewed as the incidence matrix for a bipartite graph on $n + n$ vertices, and the permanant is the number of perfect matchings in such a bipartite graph.
Note also the similarity of the definition of determinant and the permanant (both are the sum of $n!$ products). While determinant is well known to be polynomial time computable, the computation of permanant, even when restricted to the 0-1 case, if $\#P$-complete. With an easy extension of inclusion-exclusion principle, Ryser gave an algorithm that computes the permanant in $O(2^n n)$ time instead of $O(n!n)$.
For any $M \subseteq [n]$, let $r_{i,M}$ be the sum of all the elements on row i with column number not in M. Let $P(M) = \prod_{i=1}^{n} r_{i,M}$ be the sum of product of n-tuples where we select one number from each row but avoid the column numbers in M.

Fact 5.3 (Ryser 1963)**.**

$$perm(A) = \sum_{M \subseteq [n]} (-1)^{|M|} P(M).$$

Note that, nowadays a dynamic programming algorithm with similar time complexity is basic to the programming contest community. But, as Shang Jingbo pointed out to me, Ryser's algorithm has a much smaller space complexity.

Example 5.5 (Euler's totient function)**.** *$n = p_1^{a_1} p_2^{a_2} ... p_r^{a_r}$, the number of integers $1 \leq k \leq n$ s.t. $(n, k) = 1$ is*

$$\phi(n) = n - \sum_i \frac{n}{p_i} + \sum_{i<j} \frac{n}{p_i p_j} - ... = n \prod_i (1 - \frac{1}{p_i}). \tag{5}$$

Proof. Here $E_i = \{j : p_i | j\}$. □

Theorem 5.4 (Euler 1789)**.** *For $(a, n) = 1$, $a^{\phi(n)} \equiv 1 \mod n$.*

This is left as an easy exercise in algebra.

Theorem 5.5.

$$\sum_{d|n} \phi(d) = n.$$

Proof. Partition $[n]$ into the g.c.d. with n. □

Definition 5.2 (The Möbius function)**.**

$$\mu(n) = \begin{cases} 1 & \text{if } n \text{ is the product of an even number of distinct primes} \\ -1 & \text{if } n \text{ is the product of an odd number of distinct primes} \\ 0 & \text{if } n \text{ is not square-free} \end{cases}$$

Note that $\mu(n) = \mu(1, n)$ as we defnied on poset $([n], |)$. So the following is just a restatement of something we already know.

Fact 5.6. *For $n \in \mathbf{N}$,*

$$\sum_{d|n} \mu(d) = \chi_{n=1}$$

And the Möbius inversion applied on the divisibility poset, we have

Theorem 5.7 (Möbius inversion formula)**.** *Let f and g be two functions on $\mathbf{N}$. If*

$$f(n) = \sum_{d|n} g(d),$$

then

$$g(n) = \sum_{d|n} \mu(d) f(\frac{n}{d}) = \sum_{d|n} \mu(\frac{n}{d}) f(d).$$

And vice versa.

Here is a double counting proof, without using the general Möbius inversion.

Proof. Start from the r.h.s.,

$$\begin{aligned}\sum_{d|n} \mu(d) f(\frac{n}{d}) &= \sum_{d|n} \mu(d) \sum_{t|(n/d)} g(t) \\ &= \sum_{t|n} g(t) \sum_{d|(n/t)} \mu(d)\end{aligned}$$

By Fact 5.6, the last term is non-zero only if $t = n$. □

Note that the (middle part of the) formula for the Euler totient function (5) can be written as

$$\phi(n) = \sum_{d|n} \mu(d) \frac{n}{d}.$$

Use Möbius inversion we confirm Theorem 5.5.

Example 5.6. *Necklace of length n, with c colours, same configuration can be rotated.*

Solution. c^n is clearly an overcount of the answer. But we can give it an combinatorial explanation as $c^n = \sum_{d|n} d \cdot g(d)$, where $g(d)$ is defined as the number of necklace of length d and no period smaller than d. Use the Möbius inversion,

$$dg(d) = \sum_{t|d} \mu(\frac{d}{t}) c^t.$$

Also from the definition of g, out answer is $\sum_{d|n} g(d)$. So we get

$$\begin{aligned}\sum_{d|n} g(d) &= \sum_{d|n} \sum_{t|d} \frac{1}{d} \mu(\frac{d}{t}) c^t \\ &= \sum_{t|n} c^t \frac{1}{n} \left(\sum_{d'|n/t} \frac{n/t}{d'} \mu(d') \right) \\ &= \frac{1}{n} \sum_{t|n} \phi(\frac{n}{t}) c^t\end{aligned}$$

□

As one more exercise on couting by rows and columns, one can show

Lemma 5.8 (Cauchy, Frobenius, but not Burnside's)**.** *Let G be a permutation group acting on a set X. For $g \in G$ let $\psi(g)$ denote the number of fixed points for g, then the number of orbits of G is equal to $\frac{1}{|G|}\sum_g \psi(g)$.*

Exercise 5.4. *Use the lemma to solve the necklace problem again.*

Example 5.7. *Count the number of monic irreducible polynomials of degree n over the field $\mathbf{F}_q$ ($q = p^t$ for some prime p and integer t).*

Solution. List all the m.i.p's $f_1, f_2, ..., f_T, ...$ Let d_i be the degree of f_i. And N_d be the number of occurrence of $d_i = d$.
There are q^n monic polynomials over $\mathbf{F}_q$ in total.
Key point: By the unique factorization, every monic polynomial is uniquely expressed as

$$f(x) = f_1(x)^{a_1} f_2(x)^{a_2} ... f_T(x)^{a_T} ...,$$

we map f to the term $z^{a_1 d_1} z^{a_2 d_2} ....$ (One thing I did bad in class was to use both x for the variable in the polynomial and the placeholders in the g.f.) The following clothlines are equal

$$\sum_0^\infty q^n z^n = (1 + z^{d_1} + z^{2d_1} + z^{3d_1} ...)(1 + z^{d_2} + z^{2d_2} + z^{3d_2} ...)...$$

i.e.

$$\frac{1}{1-qz} = \prod_i \frac{1}{1-z^{d_i}} = \prod_d \left(\frac{1}{1-z^d}\right)^{N_d}$$

Taking log on both sides.

$$\sum_{n\geq 1} \frac{q^n}{n} z^n = \sum_d N_d \sum_{j\geq 1} z^{jd}/j.$$

Compare the coefficients for z^n we get

$$q^n = \sum_{d|n} dN_d.$$

By Möbius inversion, we have

$$N_n = \frac{1}{n}\sum_{d|n} \mu(\frac{n}{d}) q^d.$$

□

Note: Consider the term with smallest power of p, we see that the above N_n is always in the form $p^{smallest}(1 + pX)$, which is positive. When $q = p$ is a prime, this means there is at least one monic irriducible polynomial over $\mathbf{F}_p$ of degree n, and implies the existence of a field of size p^n for every n.

6 Stirling numbers

Definition 6.1 (Stirling numbers of the 2nd kind and the Bell number)**.** *The Stirling numbers of the 2nd kind is* $S(n,k) :=$ *the number of partitions of* $[n]$ *into exactly* k *parts.* $S(0,0) := 1$ *and* $S(n,k) := 0$ *if* $n \leq 0$ *or* $k \leq 0$ *but* $(n,k) \neq (0,0)$*.*
The Bell number for n *is the total number of partitions of* $[n]$*:*

$$B(n) := \sum_{0}^{n} S(n,k).$$

The first values are

$n \to$	0	1	2	3	4	5
0	1	0	0	0	0	0
1	0	1	1	1	1	1
2	0	0	1	3	7	15
3	0	0	0	1	6	25
4	0	0	0	0	1	10
5	0	0	0	0	0	1

Fact 6.1.

$$S(n,1) = S(n,n) = 1, S(n,2) = 2^{n-1} - 1, S(n,n-1) = \binom{n}{2},$$

$$S(n,k) = S(n-1,k-1) + kS(n-1,k).$$

Exercise 6.1. *For any fixed* k*,* $S(n,k) \in \Theta(k^n)$

By counting the number of surjective mappings from $[n]$ to $[k]$, and using PIE, we get

Fact 6.2.

$$k!S(n,k) = k^n - \binom{k}{1}(k-1)^n + \binom{k}{2}(k-2)^n \cdots + (-1)^k \binom{k}{k} 0^n. \qquad (6)$$

Note that the combinatorial proof works when $n < k$, in which case we proved that the r.h.s. is 0.
And count the mappings from $[n]$ to $[x]$, we get

Fact 6.3.

$$x^n = \sum_{k=0}^{n} \binom{x}{k} k! S(n,k) = \sum_{0}^{n} (x)_k S(n,k).$$

Let $f_k(x)$ be the g.f. for the k-th row, use the recurrence we get

$$x(f_{k-1}(x) + kf_k(x)) = f_k(x),$$

Note that $f_0(x) = 1$, so

Fact 6.4. *The g.f. for the k-row of $S(n,k)$ is*

$$f_k(x) = \frac{x}{1-kx} f_{k-1}(x) = ... = x^k \prod_{t=0}^{k} \frac{1}{1-tx}.$$

Exercise 6.2. *From the g.f. we find $S(n,k)$ is the coefficient of x^{n-k} in*

$$\prod_{t=0}^{k} (1 + tx + t^2x^2 + t^3x^3 + ... + t^nx^n).$$

Give a combinatorial proof for this.

We do some interpolation

$$g(x) = \prod_{t=0}^{k} \frac{1}{1-tx} = \frac{A_1}{1-x} + \frac{A_2}{1-2x} + ... + \frac{A_k}{1-kx}$$

Multiply both sides by $1-tx$ and use $x = 1/t$, we get a simple form for A_t, and $S(n,k) = \mathcal{C}_{n-k} g = A_1 + A_2 2^{n-k} + A_3 3^{n-k} + ... + A_k k^{n-k}$. We get (6) again.
Let $F_k(x)$ be the e.g.f. for the k-th row, just write out (6) again, we have

Fact 6.5. *The e.g.f. for the k-row of $S(n,k)$ is*

$$F_k(x) = \frac{(e^x - 1)^k}{k!}$$

Exercise 6.3. *Find a combinatorial proof for*

$$B(n+1) = \sum_{i=0}^{n} \binom{n}{i} B(i)$$

and solve the e.g.f. for $B(n)$.

Definition 6.2 (Stirling numbers of the 1st kind)**.** *The signless Stirling numbers of the 1st kind is* $c(n,k) :=$ *the number of permutations of* $[n]$ *with exactly* k *cycles.* $c(0,0) := 1$ *and* $c(n,k) := 0$ *if* $n \leq 0$ *or* $k \leq 0$ *but* $(n,k) \neq (0,0)$.
The Sirling numbers of the first kind is

$$s(n,k) := (-1)^{n-k} c(n,k).$$

Obviously, $\sum_{k=0}^{n} c(n,k) = n!$. Try to fill in the first values of the $c(n,k)$ table results in

$n \to$	0	1	2	3	4	5	6	7
0	1	0	0	0	0	0	0	0
1	0	1	1	2	6	24	120	720
2	0	0	1	3	11	50	274	.
3	0	0	0	1	6	35	225	.
4	0	0	0	0	1	10	85	.
5	0	0	0	0	0	1	15	.
6	0	0	0	0	0	0	1	.

Here is a good fact you need while doing the labor.

Fact 6.6.

$$c(n,k) = c(n-1,k-1) + (n-1)c(n-1,k).$$

For the proof, consider whether n is in a singleton cycle or not.

Theorem 6.7. *For* $n \geq 0$*, we have*

$$\sum_{k=0}^{n} c(n,k)x^k = x(x+1)\cdots(x+n-1)$$

and

$$\sum_{k=0}^{n} s(n,k)x^k = (x)_n.$$

Proof. For the first part, we present two proofs.
Combinatorial: View both sides as polynomials in x, and let x be any positive integer, show that both sides are counting the same thing. The l.h.s is clear, and for the r.h.s., consider we put the numbers one by one, each new number can start a cycle with one of the x colours, or can be inserted into an existing cycle after any of the previous elements.

Inductive: Essentially the same as in textbook. Use the recursive relation for $c(n,k)$.
The second part easily follows. $\square$

Now we put two things together

$$\sum_{k=0}^{n} s(n,k)x^k = (x)_n, \quad \sum_{k=0}^{n} S(n,k)(x)_k = x^n.$$

Let us fix some N, and consider $\mathcal{P}$, the polynomials of degree less than N over a field form a vector space of dimension N. ${x^k}_{k=0}^{n-1}$ is a basis for the vector space, so is ${(x)_k}_{k=0}^{n}$. And the pair above asserts that $S(n,k)$ and $s(n,k)$ are the transformations between the bases, and the matrics for $S(n,k)$ and $s(n,k)$ inverts each other.
You may want to check this from our initial tables for S and c, and remember to put $(-1)^{n-k}$ in the c-table to get the s values.

7 Integer partitions

The number 6 can be expressed as the sum of positive integers in various ways. e.g., $1+3+2$, $1+1+1+1+1+1$, 6, $2+3+1$, $4+2$, etc. When the order of the parts does not matter, we focus only on the partitions with non-increasing parts, e.g. $3+2+1$, $4+2$, $2+2+2$, etc. While the number of ordered partitions is quite easy to obtain, many interesting topics sprung from the study of unordered partitions.

Exercise 7.1. *There are 2^{n-1} ordered partitions of n.*

Definition 7.1. *Define $p(n)$ to be the number of unordered partitions of n. i.e. $p(n)$ is the number of integer solutions to the equation*

$$n = x_1 + x_2 + ... + x_n, \quad x_1 \geq x_2 \geq ... \geq x_n \geq 0.$$

Or, the number of integer solutions to the equation

$$n = 1y_1 + 2y_2 + ... + ny_n, \quad y_i \geq 0.$$

The last view gives a natural bijection between the integer partitions and the terms in the expansion of

$$(1 + x + x^2 + x^3 + ...)(1 + x^2 + x^4 + x^6 + ...)...(1 + x^k + x^{2k} + ...)...$$

Thus,

Fact 7.1. *The generating function for the partition function is*

$$\sum_{n\geq 0} p(n)x^n = \prod_{k\geq 1} \frac{1}{1-x^k}.$$

It is often beautiful and useful to draw diagrams of dots or squares for integer partitions, for example, the partition $5+5+4+2+2$ is drawn in Figure2. The picture on the left is called *Ferrers diagram*; the picture on the right is often called *Young diagram* or *Ferrers board*.
Turn your head in the right angle, when we interchange the rows and columns of the Ferrers diagram, we get its *conjugate*.
If a partition of n has k parts, then its conjugate has k as its largest number. And vice versa. So

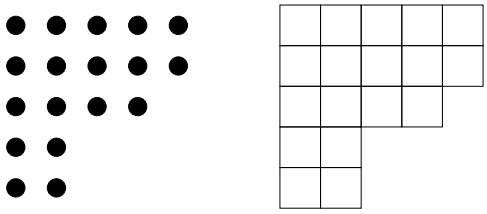

Figure 2: The Ferrers diagram and Young diagram.

Example 7.1. *The number of partitions of n where the largest part is k, equals the number of partitions of n into exactly k parts.*

Example 7.2. *For $n \geq k$, the number of partitions of n into k parts, equals the number of partitions of $n-k$ into parts that are at most k.*

Example 7.3. *The number of partitions of n where all parts are distinct equals the number of partitions of n where the smallest part is 1, and adjacent parts differ by at most 1.*

The following example is even graphical in its statement.

Example 7.4. *A partition is* self-conjugate *if it equals its conjugate. Prove that the number of self-conjugate partitions of n equals the number of partitions of n into unequal odd parts.*

Theorem 7.2. *The number of partitions of n into odd parts equals the number of partitions of n into distinct parts.*

Proof. (by generating function) Compare the generating functions for both sides.

$$\sum_{k\geq 1}(1+x^k) = \sum_{k\geq 1}\frac{1-x^{2k}}{1-x^k} = \sum_{k \text{ is odd}}\frac{1}{1-x^k}$$

□

Proof. (combinatorial) Given a partition with odd parts, for a give odd part $2k+1$ which appear $t = 2^{t_1} + 2^{t_2} + ... + 2^{t_k}$ times, where t_i's corresponding to the 1's in the unique binary representation of t. We replace them by the parts $2^{t_i}(2k+1)$. Check this mapping is invertible by showing the explicit inversion. □

The numbers 1, $5 = 1 + 4$, $12 = 1 + 4 + 7$, $22 = 1 + 4 + 7 + 10$, ... $\omega(k) = (3k^2 - k)/2$ are called *pentagonal numbers* because they can be drawn that way (Figure3). We also consider $\omega(-k) = (3k^2 + k)/2 = \omega(k) + k$.

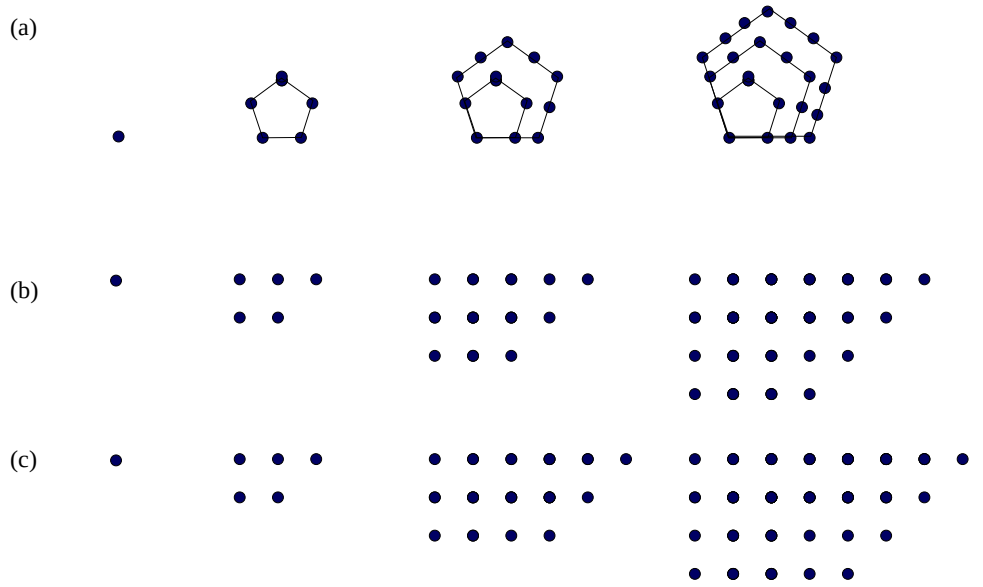

Figure 3: (a) the pentagonal numbers; (b) the pentagonal numbers $\omega(k)$; (c) the pentagonal numbers $\omega(-k)$.

We denote $p_o(n)$ (resp. $p_e(n)$) to be the number of partitions of n into odd (resp. even) number of distinct parts.
Consider the generating function $\mathcal{P}(x) = \prod_{n\geq 0} \frac{1}{1-x^k}$, its inverse is

$$\mathcal{P}^{-1}(x) = \prod_{n\geq 0}(1 - x^k) = \sum_{n\geq 0}(p_e(n) - p_o(n))x^n.$$

Euler first discovered that $p_e(n) = p_o(n)$ unless n is $\omega(\pm k)$ for some k.

Theorem 7.3 (Euler's pentgonal theorem, 1740s)**.**

$$p_e(n) - p_o(n) = \begin{cases} (-1)^t & \text{if } n = \omega(t) \text{ or } n = \omega(-t) \\ 0 & \text{otherwise} \end{cases}$$

Proof. (Franklin 1881, sketch) Draw the Ferrers diagram, call the bottom row the *base*, and the 45 degree run from the top-right corner the *slope*. Consider the operation ϕ to transform a partition into another:

```
if base is no more than the slope:
    take it off and make it the new slope,
    to the right of the old slope.
if base if longer than the slope:
    take off the slope and make it a new row below the base.
```

It is easy to see the mapping ϕ is (almost) an involution and pairs the partitions with odd number of distinct parts and the partitions with even number of distinct parts.
The only place where ϕ cannot be defined are exactly those Ferrers diagram with polygonal numbers as shown in Figure3 (b) and (c).

□

8 Local observers and global energy

I always like to introduce advanced students (advanced meaning the students know how to write a sorting program) to the theory of discrete structures by the following three seemingly similar problems. All three involves the word *sort*, yet none of the solutions requires actually doing the sorting. Their answers are quite different; each one is related to a very different problem; yet the justifications of the answers become, again, quite similar. That is what I want to talk about in this section.
We introduce the examples in terms of permutations. It is easy to see they work for general array of distinct numbers with slight modifications.

Example 8.1. *Given a permutation a of $[n]$. In each step you are allowed to swap two adjacent elements in the array. For any such array, how can you decide the minimum number of steps one needs to sort it?*

Solution. The *bad pairs* are defined to be the pairs (i, j) such that $i < j$ and $a_i > a_j$. The minimum number of swaps one needs is exactly the number of bad pairs.
To justify. For any state S, define $E(S)$ to be the number of its bad pairs. Observe how E changes when we swap two adjaent elements. □
This also proves, no matter we do insertion sort, or bubble sort, as long as we swap one bad pair in each step, we achieve the minimum number of swaps.

Example 8.2. *Given a permutation a of $[n]$. In each step you are allowed to take out one elements and insert it anywhere in the array. For any such array, how can you decide the minimum number of steps one needs to sort it?*

Solution. Now the answer is $n - E(S)$, where $E(S)$ is defined to be the length of the longest increasing subsequence of the array.
To justify, observe how $E(S)$ can change in one step. □

Example 8.3. *Given a permutation a of $[n]$. In each step you are allowed to swap two elements (not necessary adjacent) in the array. For any such array, how can you decide the minimum number of steps one needs to sort it?*

Solution. The answer is $n - E(S)$, where $E(S)$ is defined to be the number of cycles in the permutation. □

This also work backwards very nicely. As my friend Bai Haoquan suggested, to compute the number of cycles, you do not need to do the obvious depth first search. You just need to go from 1 to n, and fix each number as necessary. By doing this one always increases the number of cycles by 1.

Example 8.4. *Let $\mathcal{L} = (X, \preccurlyeq)$ be a finite lattice. For any $X_1, X_2 \subseteq X$, define*

$$\mu(X_1, X_2) = \sum_{x_1 \in X_1, x_2 \in X_2} \mu(x_1, x_2).$$

Suppose A, B, and C is a partition of X such that A is an ideal and C is a filter. Prove that

$$\mu(A, C) = \mu(B, B) - 1.$$

You can solve this problem by just sit there and looking. The statement itself suggests an invariance. You just need to prove no matter how the state (here the partitions A, B, and C) changes,

$$E(A, B, C) := \mu(A, C) - \mu(B, B)$$

does not change. (Note that the value is very easy to compute for some special states.) The nice steps will be moving one minimal element from C to B, etc.

We restate Example 1.4. The game attributed to Maxim Kontsevich, dated 1981.

Example 8.5. *In the beginning there is a peg on the position $(0, 0)$. In each step one may remove a peg at (a, b) and put one peg on each of $(a + 1, b)$ and $(a, b+1)$, provided both positions are empty. How can you play the game so that there will be no peg in the $x + y \leq 1$ area? How about $x + y \leq 2$, $x + y \leq 3$, etc.?*

Solution. For each integer point (x, y), define $\phi(x, y) := 2^{-x-y}$, and for each state S of pegs, define

$$E(S) := \sum_{x,y:\text{there is a peg on } (x,y)} \phi(x, y).$$

Observe that $E(S)$ does not change in each step. □

In the same line, we have a more famous and older problem, due to J. H. Conway, dated 1961.

Example 8.6 (Conway's soldiers)**.** *Below or on the horizontal line* $y = 0$ *you may place as many pegs as you like, with one soldier on a distinct integer point. Then you start the game, in each step you can do a solitaire jump. Prove that it is impossible for any peg to reach* $(0, 5)$*.*

I complete the short interlude with one of my favorite puzzles. Depending on whether you are lucky or whether you have the right glasses, this may take you one second, or a week to solve.

Example 8.7. *On an* $n \times n$ *chessboard you can pick* $n - 1$ *squares to colour yellow. Then in each step, whenever there is a square has two or more yellow neighbours, it becomes yellow itself.*
Prove that no matter how you choose the initial $n-1$ *squares, it is impossible to make the whole board yellow in the end.*

9 Graphs

First paper on graph theory: Euler 1736 about the bridges in Königsberg. The first book on graph theory appeared exactly 200 years later in 1936, by the mathematician in a quite different place but named König.
In most cases, a graph is just a relation on a set. It abstracts a set of objects and their pairwise relations.
We define a *graph* G as an ordered pair (V, E) with an incidence function. Unless otherwise stated, V and E are finite.
For *simple graphs* (graphs without parallel edges and without self-loops), we do not need the incidence function and view E as a subset of $\binom{V}{2}$. In our class, unless specifically mentioned, the term graph refers to simple graphs.

Definition 9.1. *A* simple graph *is a pair $G = (V, E)$, where V is a finite set (usually called the vertex set) and $E \subseteq \binom{V}{2}$.*

For *directed graphs* (a.k.a. *digraph*), E is viewed as a subset of $V \times V$. In set theory, this is just a relation on V.

Definition 9.2. *A* directed graph *is a pair $G = (V, E)$, where V is a finite set and $E \subseteq V \times V$.*

It is standard to write $v(G) := |V|$ and $e(G) := |E|$. It is also conventional to write them as n and m. They are called the *order* and *size* of the graph, respectively.
We write $uv \in E$ (or $u \sim v$) to indicate that there is an edge between u and v in G. In this case u and v are said to be *adjacent*, they are *incident* to the edge. For directed graphs, we write $\overrightarrow{uv} \in E$, or $u \to v$. u is called the tail of the edge, and v the head.
An *oriented graph* is a simple graph where we pick a direction for each edge. Clearly each simple graph can be oriented in 2^m ways. It can also be thought of the class of directed graphs where $\overrightarrow{uv}$ and $\overrightarrow{vu}$ cannot both present. i.e., E is an asymmetric relation on V.
It is often nice to view a graph as a directed graph where edges come in pairs. i.e., E is a symmetric relation (and maybe nowhere reflexive).

Example 9.1. *There are $2^{\binom{n}{2}}$ (simple) graphs on $[n]$, 2^{n^2} directed graphs on $[n]$, and $3^{\binom{n}{2}}$ oriented graphs on $[n]$. Try to view each of these from different angles.*

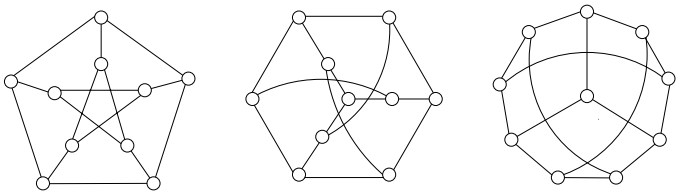

Figure 4: The Petersen graph.

A *drawing* of a graph maps V to a set of n distinct points P in the plain, each edge mapped to a curve between its end points, and any point in P does not lie on the interior of such curves. A graph has infinitely many drawings; they can look quite different.
In a graph, the *neighbours* of the vertex v is the set $N(v) := \{u : uv \in E\}$. The *degree* of v is $d(v) := |N(v)|$. The smallest (resp. the biggest) degree in G is denoted $\delta(G)$ (resp. $\Delta(G)$). In a digraph, N^-, N^+, d^-, d^+, δ^-, δ^+, Δ^-, Δ^+ are similarly defined, where $-$ means edges going into a vertex, and $+$ means edges going out.
By counting the "number of vertex v incident to edge e" relations in two ways, we get the following theorem, which is often the first theorem one learns in graph theory.

Theorem 9.1. $\sum_v d(v) = 2m$. *As a result, in any graph the number of vertices with odd degree is even.*

We look at some special examples of graphs along with more definitions.

Example 9.2 (Erdös number)**.** V = *all the* mathematicians, $E = \{xy :$ *x and y has a joint paper*$\}$. *The* Erdös number *of x is defined as the distance from x to Erdös in this graph.*

Example 9.3 (Petersen graph)**.** *Pictures in Figure4 are the drawings of the same* Petersen graph.

To prove the three drawings in Figure4 are actually the same graph, it is nice to know another construction of Petersen graph and then show all three drawings satisfy the construction.

Fact 9.2. *The Petersen graph can be defined as (V, E), where $V = \binom{[5]}{2}$, and (we simply write ab for the set $\{a, b\}$) $ab \sim cd$ iff a, b, c, d are all distinct.*

The *complement* of $G = (V, E)$, denoted $\overline{G}$, is $(V, \binom{V}{2} - E)$. It is sometimes nice to view the edges as red and non-edges as blue to feel the two complementary graphs as more symmetric.
A *walk* of length k in a graph is a sequence of vertices $u_0u_1...u_k$ where u_iu_{i+1} is an edge. If all the u_i's are different, it is called a *path*. If $u_0 = u_k$, it is called a *closed walk*. If all the vertices are different except $u_0 = u_k$, it is called a *cycle*. P_n is the graph with n vertices which form an $(n-1)$-path. C_n is the graph on n vertices which form an n-cycle.
A *complete graph* of order n is K_n where all the $\binom{n}{2}$ possible edges are present. Its complement is the *empty graph of order* n. An orientation for the complete graph is called a *tournament*. For a labeled graph with n vertices, there are $2^{\binom{n}{2}}$ tournaments, they can be viewed as all the possible outcomes of a round robin tournament with n teams.

Example 9.4. *The complement of P_4 is again an P_4, the complement of C_5 is also a C_5.*

A graph is *connected* if there is a walk (therefore there is a path) between each pair of vertices. Otherwise it is *disconnected*, in which case we can find a partition of $V = A + B$ such that there are no edges between A and B.

Fact 9.3. *If G is disconnected, then we can find a partition $V = A + B$ such that there are no edges between A and B.*

Proof. (Sketch) Fix a vertex v, let A be the set of vertices reachable by v, and $B = V - A$. □

Example 9.5. *If G is not connected, then there are at most $\binom{n-1}{2}$ edges in G.*

Proof. (Sketch) Let $V = A + B$ where there are no edges between A and B, w.l.o.g, suppose $|A| \leq |B|$. There are at least $|A| \cdot |B|$ pairs that cannot be in E (and we can in fact add all the other pairs into E). Discuss when $|A||B|$ is minimized. (One way to do this is shifting $|A|$.) □

If G is not connected, we find the partition of vertices $V = A + B$ where $\overline{G}$ contains all the edges between A and B, and thus connected. So,

Fact 9.4. *Either G or its complement is connected.*

Definition 9.3. *A closed path in a graph using every edge exactly once is called an* Eulerian circuit*, and a graph that has such a path is called an* Eulerian graph*.*

And here is usually the second theorem every one knows about graph theory.

Theorem 9.5. *A finite graph G (with possible parallel edges and) with no isolated vertex is Eulerian iff it is connected and each vertex has even degree.*

Definition 9.4. *A path (reps. cycle) in a graph touching every vertex exactly once is called a* Hamiltonian path (resp. cycle)*. A graph is called* Hamiltonian *if it has a Hamiltonian cycle.*

As we see, E is easy, H is hard. Deciding Hamiltonian property is NP-complete.

Exercise 9.1. *The Petersen graph is not Hamiltonian.*

Theorem 9.6 (Dirac 1952)**.** *If G is a graph on $n \geq 3$ vertices and $\delta(G) \geq n/2$, then G is Hamiltonian.*

which can be clearly derived from

Theorem 9.7 (Ore 1960)**.** *If G is a graph on $n \geq 3$ vertices and $d_u + d_v \geq n$ whenever $u \not\sim v$, then G is Hamiltonian.*

which can be clearly derived from

Theorem 9.8 (Bondy-Chvátal 1972)**.** *Let $G = (V, E)$ be a graph, $u, v \in V$ and $d_u + d_v \geq n$. Let $G' = (V, E \cup \{uv\})$. Then G is Hamiltonian iff G' is.*

Proof. If G has a Hamiltonian cycle, certainly G' does too.
If G' is Hamiltonian, let C be a Hamiltonian cycle in G'. If C does not use the edge uv, then C is also a Hamiltonian cycle in G. Otherwise, suppose $C = (u, v, x_1, x_2, ..., x_{n-2})$. Note that in G $u \not\sim v$ and $d_u + d_v \geq n$. Suppse $x_{a_0}, x_{a_1}, ..., x_{a_k}$ are neighbours of v, where $a_0 = 1$. If u is not adjacent to any of x_{a_1-1}, ..., x_{a_k-1}, then d_u is at most $n - 2 - k$, and $d_u + d_v < n$, a contradiction. So there is $i = a_s$ such that $v \sim x_i$ and $u \sim x_{i-1}$, and

$$(v, x_i, x_{i+1}, ..., x_{n-2}, u, x_{i-1}, x_{i-2}, ..., x_1)$$

is a Hamiltonian cycle in G. □

A *tree* is a connected graph that contains no cycles. It is left as an exercise to verify that in a graph 2 of the following 3 conditions imply the 3rd one: (a) G is connected, (b) G has no cycle, (c) G has exactly $n-1$ edges.
A graph without cycles is called a *forest*. Each connected component of a forest is a tree. (Note, an isolated point is a tree of order 1.)
A vertex with degree 1 in a tree is call a *leaf*. By counting the degrees, we have

Fact 9.9. *In a tree with order $n \geq 2$, there are at least two leaves.*

Exercise 9.2. *An oriented graph is called* balanced *if for each vertex v,*

$$|d^+(v) - d^-(v)| \leq 1.$$

Prove that, every simple graph G has a balanced orientation.

Proof. (sketch) Induction on m. Suppose the statement holds for all simple graphs with less edges than G.
If there is a cycle C in G, Take a balanced orientation of $G - C$ and orient C in one order. Otherwise G is a forest. Take any tree in G, there are two degree 1 vertices x and y. Find a path P from x to y. Take a balanced orientation of $G - P$ and orient P in one direction. □

Another very nice proof is provided by two students in an exam in 2011.

Proof. (Qu Jun and Liu Sizhuang) We may assume the graph is connected. Because the number of vertices with odd degree is even, say $2t$, we arbitarily pair them, and add one (magic) edge to each pair. Thus get a graph G' which is Eulerian, orient the edges according to a Eulerian tour, so every vertex has the same in-degree and out-degree. Now remove the magic edges to get an orientation of G, each vertex has at most one edge removed. □

In 1889 Cayley published his famous formula without a proof.

Theorem 9.10 (Cayley 1889)**.** *There are n^{n-2} different spanning trees in K_n on $[n]$.*

Note that here we do not consider isomorphic trees. Two isomorphic trees with different vertex labelings are considered different.

Proof. (Prüfer code) Given any labeled tree, we view it as rooted on n. We remove the leaves of the tree one by one. In step i ($1 \leq i \leq n-1$), let a_i be the smallest leaf and b_i be its neighbour.
b_{n-1} is always n. $(b_1, b_2, ..., b_{n-2})$ is a vector in $[n]^{n-2}$, called the *Prüfer code* of the tree. We claim that the code can take any value in $[n]^{n-2}$, and two different trees have different codes. Therefore there are n^{n-2} different trees. We check that from any vector in $[n]^{n-2}$ we can uniquely decode one tree. The process is almost determined: a_i is the smallest number that does not appear in $a_1, ..., a_{i-1}, b_i, ..., b_n$. (*exercise*: check our claims.) □

It is easy to see that any vertex v appears in the Prüfer code exactly $d_v - 1$ times. Since the Prüfer code maps 1-1 between trees and $[n]^{n-2}$, given the degrees d_1, d_2, ..., d_n, there are

$$\binom{n-2}{d_1-1, d_2-1, ..., d_n-1} = \frac{(n-2)!}{(d_1-1)!(d_2-1)!...(d_n-1)!}$$

labeld trees on $[n]$ with the given degrees.
A graph is *bipartite* if V can be partitioned into A and B such that all the edges are between A and B. It is left as an exercise to verify that a graph is bipartite iff it contains no odd cycles. And if the graph has k connected components, it has 2^k different bipartitions. The *complete bipartite graph* $K_{m,n}$ is the graph with m points in one part, n points in the other, and all the mn possible edges. $K_{1,n}$ is called a star.
A *subgraph* of G is a subset of edges along with all the supporting vertices. Let $V' \subseteq V$, the *induced subgraph* of G, denoted $G|_{V'}$ or $G[V']$ has vertex set V' and edge set $E' = \{uv : u, v \in V', uv \in E\}$.
We can define isomorphic graphs as those who can be drawn in the same way, i.e., who have the same drawing. Formally, as we always do for algebraic strctures:

Definition 9.5. *H is* isomorphic *to G ($H \cong G$) if there is a bijection f between $V(H)$ and $V(G)$ such that uv is an edge in $E(H)$ iff $f(u)f(v)$ is an edge in $E(G)$.*

For digraphs, we may use $\overrightarrow{uv} \in E(H)$ iff $\overrightarrow{f(u)f(v)} \in E(G)$.
The isomorphic relation is an equivalence relation thus it gives a partition of all the graphs. Graphs isomorphic to each other share the same graph properties, they must share the same drawing, have the same degree sequence, the

same connectivity, the same chromatic number, the same girth, etc. (These are obvious and tedious to write down. So, except for the example below, it will be good enough if you check these silently.) When we talk about a specific graph, e.g., P_4, K_5, we usually think about the whole equivalence class. We also say H is a subgraph of G if it is isomorphic to a subgraph of G.

Example 9.6. *If H and G are isomorphic, and G is connected, then H is connected.*

Proof. Let f be an isomorphism from H to G. For any non-empty partition $V(H) = A + B$, $f(A)$ and $f(B)$ is a non-empty partition of $V(G)$. Since G is connected, there exists $a \in A$ and $b \in B$ s.t. $f(a)f(b) \in E(G)$. And since f is an isomorphism, $ab \in E(H)$. Thus there exists edges between any partitions of $V(H)$, therefore H is connected. □

An *automorphism* of a graph G is a isomorphism of G to itself. i.e., a permutation of V such that $uv \in E$ iff $\pi(u)\pi(v) \in E$. It is easy to check that the set of automorphisms, denoted $Aut(G)$, form a subgroup of the symmetric group on V.

Exercise 9.3. *Find $Aut(G)$ for $G =$ (a) K_n, (b) P_n, (c) C_n, (d) Petersen graph.*

Example 9.7. *There are no graphs on* [6] *such that $|Aut(G)| = 1000$ or $|Aut(G)| = 42$.*

Exercise 9.4. *For any positive integer n, find a graph G with $|Aut(G)| = n$.*

A graph is *vertex transitive* if for any two vertices u and v, there is an automorphism that maps u to v.
A graph is *k-regular* if every degree is k, K_n is $(n-1)$-regular, the Peterson graph is 3-regular. 1-regular graphs are called *matchings*. 2-regular graphs are the union of disjoint cycles. 3-regular graphs can actually be quite wild.
A digraph is *k-diregular* if $d^+(v) = d^-(v) = k$ for all v.

Example 9.8. *A digraph is vertex transitive, then it is k-diregular for some k. Also show examples that the converse is not true.*

Proof. If the graph is vertex transitive, for any two points u and v, there is an automorphism mapping u to v, and it is a bijection between the in(out)-neighbours of u and the in(out)-neighbours of v. So $d^+(u) = d^+(v)$, and $d^-(u) = d^-(v)$.
Hence, $d^+(u) = k_1$ and $d^-(u) = k_2$ for all the vertices. Also notice that $nk_1 = \sum d^+(u) = m = \sum d^-(u) = nk_2$, so $k_1 = k_2$.
For the converse, consider the union of two disjoint directed cycles of different sizes. □

Example 9.9 (The n dimensional cube, a.k.a. the boolean lattice)**.** *$V = 2^{[n]}$, A and B has an edge if the symmetric difference of A and B is of size 1. i.e., $A = B + \{x\}$ for some x or $B = A + \{x\}$ for some x.*

The cube can be drawn recursively by adding elements of $[n]$ one by one. Each step represent one direction. It can also be viewed as the graph on 2^n 0-1 vectors where two vectors are adjacent if their Hamming distance is 1. In this view it is clear that the cube is vertex transitive, n-regular, and bipartite.

Exercise 9.5. *Show that the cube is Hamiltonian.*
More general: Define a graph on $[t]^n$ where two tuples are adjacent iff they differ on only one coordinate. Show that such a graph is Hamiltonian by actually construct a circuit inductively.

Given a set of points on the cube (a collection of subsets of $[n]$), the *projection* along the i-th direction is the operation of deleting i from each set in the collection.

Theorem 9.11. *If A_1, A_2, ..., A_n are n distinct subsets of $[n]$, then we can find $x \in [n]$ such that $A_i - \{x\}$ are all distinct.*

Proof. Draw a graph on the A_i's where A_iA_j is an edge and put a label x on it if the symmetric difference of A_i and A_j is x. This means x is a bad direction.
Now, as long as there is a cycle in G, start from A_i and comes back to A_i, each label on the cycle must appear an even number of times (why?). Therefore, we can delete half of the labels while keeping the total number of different labels in the graph unchanged.
Since there are finitely many edges to start with, we must end up with a graph with no cycles. So there are at most $n - 1$ bad directions. □

Example 9.10. *3 cannibals and 3 missionaries need to cross a river with one boat. The boat can only hold two people at a time. We sincerely hope the cannibals never outnumber the missionaries on either side of the river. How do they cross the river?*

The solution uses a digraph. Each vertex represents a situation, encoded as (c, m, b), the number of cannibals on the left side, the number of missionaries on the left bank, and the location of the boat. The egdes are defined by simple rule such that $u \to v$ if v is a valid state (no monks will be eaten) and from u using one boat trip we can get v. Then we just need to find any path from $(3, 3, L)$ to $(0, 0, R)$.
Given a graph, we can combine its vertices into groups to get a higher level view of it. The resulting new graph has many details in its vertices.

Example 9.11. *Given a graph of communication network between cities. We can group all the cities by their provinces to get a graph of the communication network between the provinces.*

The final topic in this section is *graph colouring*.

Definition 9.6. *Given a set (of points) V and a set (of colours) C, a* C-colouring of V *is a function $f : V \to C$.*

We can think of f as assigning one colour from C for each point in V. Clearly there are $|C|^{|V|}$ such colourings.

Definition 9.7. *Given a graph $G = (V, E)$ and a set (of colours) C, a* proper C-colouring of G *is a C-colouring of V such that $f(u) \neq f(v)$ whenever $u \sim v$.*

Usually we only need to know the number of colours, instead of the name of the colours. We often write $[c]$-colouring as c-colouring. We are interested in the number of proper colourings of G.

Example 9.12. *There are $x(x-1)^{n-1}$ proper x-colourings of a tree on $[n]$.*

While the number of proper colourings of some special graphs can be computed easily as in the example above, one systematical way to do this by pigeonhole principle. [1] We consider all the x^n colourings, and define the

[1] Interested students may study the partition lattice of sets and of graphs, and learn properties of the lattice, as well as the fact that the number of proper colourings can be computed by a Möbius inversion on this lattice.

"bad events" as

$$E_i = \{f : V \to [x] | \text{the } i\text{-th edge is monochromatic}\}.$$

It is easy to compute $E(M)$ for each particular M. Suppose the edges correspond to the indices in M connects V into k parts, then $E(M) = k$. Note that k is determined by M, not by x, thus we have

Definition 9.8. *Let $G = (V, E)$ be a graph, there is a degree n polynomial $P_G(x)$ such that for any positive integer x, the number of proper x-colourings of G is $P_G(x)$. P_G is called the* chromatic polynomial *of G.*

Example 9.13. *Use principle of inclusion-exclusion to compute the chromatic polynomial for any tree.*

Example 9.14. *Use principle of inclusion-exclusion to compute the chromatic polynomial for $G = ([4], \{\{1, 2\}, \{2, 3\}, \{3, 1\}, \{1, 4\}\})$.*

Exercise 9.6. *Let G be a graph on n vertices and m edges, let it chromatic polynomial be*

$$P_G(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0.$$

Then $a_n = 1$, $a_0 = 0$, $a_{n-1} = -m$, $a_{n-2} = \binom{m}{2} - t$, where t is the number of triangles in G.

We conclude the introduction to graphs by several simple looking problems that have been open for decades.

Conjecture 9.1 (Caccetta-Häggkvist 1978)**.** *In any simple digraph G with n vertices and $\delta^+(G) \geq r$, there is a cycle of length at most $\lceil n/r \rceil$.*

$r = 1$ is trivial, so is $r = n/2$. $r \leq 5$ is proved. Also proved for $r \leq \sqrt{n/2}$. From a slightly different direction, it is proven there is a cycle of length at most $n/r + 73$.
The case $r = n/3$ received much attention.

Conjecture 9.2 (Seymour's second neighbourhood conjecture)**.** *In a digraph G, define $N^{++}(v) := (\cup_{v \to u} N^+(u)) - N^+(v)$. If G is an oriented graph (no cycle of length at most 2), then there is a vertex v such that $|N^{++}(v)| \geq |N^+(v)|$.*

When G is a tournament, this is conjectured by Dean, and proved by Fisher (1996). Kaneko and Locke (2001) proved for G with $\delta^+ \leq 6$. Chen, Shen, and Yuster (2003) proved there is a v s.t. $|N^{++}(v)| \geq r|N^+(v)|$, where $2r^3 + r^2 = 1$ ($r \approx 0.657$).
Note: Seymour's conjecture implies that C-H conjecture is true for graphs where both δ^+ and δ^- are at least $n/3$.
In the homework we show that any two longest path in a connected graph share a common vertex.

Open Problem 9.1 (Gallai 1968)**.** *Is it true that in a connected graph G, any 3 longest paths have a vertex in common?*

10 The pigeonhole principle and Ramsey principle

The pigeonhole principle: If n objects are distributed into c groups, then at least one group contains at least $\lceil n/c \rceil$ objects. In another language, if the points in $[n]$ are coloured by $[c]$, then there exists $x \in [c]$ such that at least $\lceil n/c \rceil$ elements are coloured x (we call them monochromatic).
Or even in another picture (This is the same as the case $r = 1$ of Ramsey theorem), if we have numbers q_1, q_2, ..., q_s, then whenever $n > \sum q_i$, and $[n]$ is c-coloured, we can always find one colour $x \in [c]$ s.t. there are q_x elements having colour x.

Example 10.1. *Pick $n+1$ numbers from $[2n]$, then we can always find 2 among those $n+1$ which are relatively prime.*

Solution. There are always two of them adjacent. □

Example 10.2. *Pick $n+1$ numbers from $[2n]$, then we can always find 2 among those $n+1$ where one divides the other.*

Solution. Each number can be written as $(2k-1)2^m$, colour it with k. □

Exercise 10.1. *Let $(a_1, a_2, ..., a_n)$ be a sequence of integers, we can always find a segment whose sum is a multiple of n. i.e., there is $1 \leq s \leq t \leq n$ s.t. $n | \sum_{i=s}^{t} a_i$.*

Hint: Consider the $n+1$ partial sums $\sum_{i=1}^{t} a_i$, for $t = 0, 1, ..., n$.

Example 10.3. *Form a set A by picking 10 numbers from $[100]$, we can always find disjoint subsets $X, Y \subseteq A$ such that they have the same sum.*

Solution. There are 1023 non-empty subsets of A, each with sum in $[1000]$. So there are two of them with the same sum. Throw away the common elements. □
Next is one of Erdös' favorite theorem, with my own favorite proof,

Theorem 10.1 (Erdös - Szekeres 1935)**.** *Given any sequence $a_1, a_2, ..., a_{mn+1}$ of $mn+1$ distinct numbers, one can always find a sub-sequence of length $m+1$ that is increasing, or a sub-sequence of length $n+1$ that is decreasing.*
In particular, in a sequence of n^2+1 distinct numbers, one can find a monotone sub-sequence of length $n+1$.

Proof. For each $i = 1, 2, ..., mn + 1$, define $f(i) = (I_i, D_i)$, where I_i is the length of the longest increasing sub-seq begins with the i-th position, and D_i is the length of the longest decreasing sub-seq begins with the i-th position. For any $i < j$, if $a_i < a_j$, then $I_i > I_j$, otherwise $D_i > D_j$. This means all the $f(i)$'s are different. They can not only take the values on $[m] \times [n]$. □

Exercise 10.2. *For any $m, n \geq 1$, there is a permutation of $[mn]$ where there is no increasing sub-seq of length $m+1$, nor decreasing sub-seq of length $n+1$.*

Example 10.4. *Let F_n be the n-th Fibonacci number. For any $k > 0$, there is n such that F_n ends with k 0's.*

In fact, we prove the more general:

Fact 10.2. *Let F_n be the n-th Fibonacci number. For any $T > 0$, there is n such that $T|F_n$.*

Proof. (key) Let $f_n = F_n \mod T$. There will be repeated adjacent pairs (modulo T). In addition, the relation $f_{k+2} \equiv f_k + f_{k+1}$ can be used to compute f_{k+2}, but also can be used to compute f_k from f_{k+1} and f_{k+2}. [2] □

We now discuss the Ramsey theory, which is a generalization of the pigeonholes, and whose proof uses the pigeonhole principle in a recursive manner. The classical introductory problem to the Ramsey theory: In a group of 6 persons, we can find either 3 of them who know each other, or 3 none of whom knows each other. Formulated more formally (and symmetrically), we think there are two graphs Y and B who complement each other, or equivalently, the edges of K_n are coloured by yellow and blue.

Theorem 10.3. *If the edges of K_6 are coloured with yellow and blue, then there is either a yellow K_3, or a blue K_3.*

On the other hand, we can construct a colouring of the edges of K_5, $f : E(K_5) \to \{Y, B\}$, such that there is no monochromatic K_3.

[2] Xiang Ziqing pointed out in class [2012] that there are general theorems about the divisibility of Fibonacci numbers. For our problem, we outline three from the sea of facts about Fibonacci numbers: (1) For any prime number p, one of F_p, F_{p+1}, F_{p-1} is a multiple of p; (2) If $p^k|F_n$, then $p^{k+1}|F_{np}$; (3) $F_n|F_{mn}$ for any m, n. All the proofs are left as exercise. This implies that we can always find $T|F_n$ very quickly, and the number is "roughly" as big as T. Note this is a much better bound than we can read out of the pigeonhole proof.

For the similar question about K_4. Consider the vertices of K_{17} as vertices on a regular 17-gon, it is left as an exercise to colour the egdes by the arc distance so that there is no monochromatic K_4.

Fact 10.4. *If the edges of K_9 are coloured Y and B, then there is either a copy of Y K_3, or a B K_4.*

Proof. Pick any vertex v. If it is incident to 4 yellow edges vu_i $(i = 1, 2, 3, 4)$, then either there is a yellow edge among the u_i's to form a yellow K_3 with v, or the u_i's form a blue K_4. On the other hand, it v is incident to 6 blue edges, then we can find either a yellow triangle among those 6 points, or a blue triangle among them, therefore form a blue K_4 together with v.
The only other case is when each vertex is incident to 3 yellow edges and 5 blue edges. This means $9 \cdot 3 = 2 \cdot e(Y)$, which is impossible. □

Now we can easily get the following:

Theorem 10.5. *If the edges of K_{18} is coloured with Y and B, then there is a monochromatic K_4.*

And one can prove the following theorem.

Theorem 10.6. *For any y and b, there is an integer $N := r(y, b)$, such that when the edges of K_N are coloured by yellow and blue, one can always find a yellow K_y, or a blue K_b. And*

$$r(y, b) \leq r(y - 1, b) + r(y, b - 1).$$

Corollary 10.7. $r(y, b) \leq \binom{y+b-2}{y-1}$. *In particular,* $r(k, k) \leq \binom{2k-2}{k-1}$.

The proof is left as an exercise. We will prove the more general Ramsey's theorem.
Clearly $r(k, l) = r(l, k)$. E.T.S. $r(2, k) = k$. $r(3, 3) = 6$, $r(3, 4) = 9$ as we proved. $r(3, 5) = 14$, $r(3, 6) = 18$, $r(3, 7) = 23$, $r(3, 8) = 28$, $r(3, 9) = 36$. $r(4, 4) = 18$ as we proved. $r(4, 5) = 25$. These are all the Ramsey numbers we know. $r(5, 5)$ is between 43 and 49, inclusive.

Open Problem 10.1. *Find, or improve the bound for, any unknown Ramsey number.*

Definition 10.1. *A* hypergraph *is $\mathcal{H} = (V, E)$ where $E \subseteq 2^V$.*

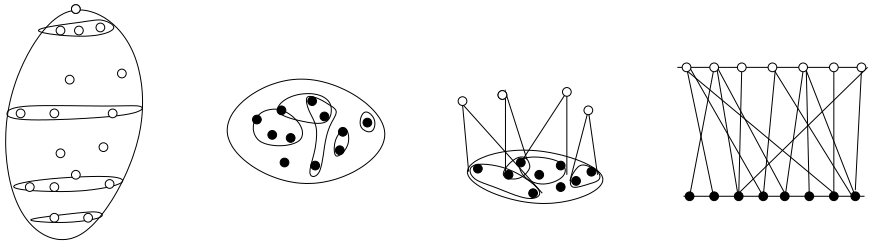

Figure 5: The different views of a hypergraph: As a set system (points in the hypercube); from above; from side with edges up in the air; as a bipartite graph. The elements are represented by black dots, the hyperedges are represented by white dots.

Note that, the only difference in the definition of a graph and a hypergraph is that each *hyperedge* can be any subset of points, vs. in a graph each edge is from $\binom{V}{2}$. If all the edges are of size r, i.e. $E \subseteq \binom{V}{r}$, the hypergraph is called *r-uniform*. So, the graphs are just 2-uniform hypergraphs.
In class we introduced several ways to draw a hypergraph. As I often claim, it matters a lot how we view combinatorial objects from the right angle. We present different ways of picturing a hypergraph in Figure5.
Now, the generic Ramsey theorem. We introduce the finite version from the paper published posthumously in 1930. Ramsey did the work in 1928, when he was 25. It roughly states, fix the number of colours s, fix the size r (this is the size of the hyperedges, or let's call them interesting small groups) and desired size q, then when n is big enough, given any s-colouring of all the interesting groups in $[n]$, we can always find q points whose all $\binom{q}{r}$ interesting goups are coloured the same.
Or as Motzkin put it: "Complete disorder is an impossiblility. Any structure will necessarily contain an orderly substructure." (I think this wishful belief is quite vague.)
Denote $K_n^{(r)}$ the *complete r-regular graph on n vertices* $([n], \binom{[n]}{r})$. In the following statement, r is the size of the hyperedge, or the size of interesting groups; s is the number of colours; q_i's are the desired sizes. Sometimes all the q_i's equal to the same desired size q. The Ramsey theorem states

Theorem 10.8 (Ramsey 1930)**.** *Let $r \geq 1$ and $q_i \geq r$ $(1 \leq i \leq s)$. There exists a minimal positive integer $N := N_s(q_1, q_2, ..., q_s; r)$ such that for any colouring $f : E(K_N^{(r)}) \to [s]$, there exists i such that there is a copy of $K_{q_i}^{(r)}$*

with colour i.

Or reformulate another way,

Theorem 10.9 (Ramsey 1930)**.** *Let $r \geq 1$ and $q_i \geq r$ $(1 \leq i \leq s)$. There exists a minimal positive integer $N := N_s(q_1, q_2, ..., q_s; r)$ such that: whenever $n \geq N$ and $\binom{[n]}{r}$ is parititioned into s families (coloured with s colours) T_1, T_2, ... T_s, one can find $1 \leq i \leq s$ and $K \subseteq [n]$ such that $|K| = q_i$ and $\binom{K}{r} \subseteq T_i$ (i.e., all the r-subsets inside K are coloured by the i-th colour).*

Here is the idea of the proof.

Proof. (idea) We use some induction. Suppose N is very big, we pick one point $v \in [N]$, and consider all the r-subsets involving v. There are $\binom{N-1}{r-1}$ such sets. We project the colours of such r-subsets to the $(r-1)$-subsets, let us call it the *elephant for* v [a name formed by Vašek in 2011]. The elephant is $\binom{[N-1]}{r-1}$ coloured with s colours, where we can use induction with s colours and $r-1$.
We want $N-1$ to be big enough, say,

$$N - 1 \geq N(q_1', q_2', ..., q_s'; r-1)$$

the q's TBD. Now set $q_1' = N(q_1 - 1, q_2, ..., q_s, r)$ (it is clear now we do induction on r first, then on $\sum q_i$). When, in the elephant, there is a q_1' set whose all $(r-1)$-subsets are of the first colour, we can find one of the following in the original system:
(1) a $q_1 - 1$ set, plus v to form a q_1 set, whose r-subsets are all coloured with the 1st colour.
(2) a q_2 set whose all r-subsets (in the original hypergraph) are coloured with the 2nd colour.
...
(s) a q_s set whose r-subsets are all coloured with the s-th colour.
Now it is obvious how to set the other q_i's. □

Note, usually how one finds the proof is not how one finally writes it. It can be an exercise to write down the above proof formally. Note that we did a double induction, for $(c, \sum q_i, r)$, the inductive hypothesis is that the theorem is correct for all $r' < r$ (no matter how big are the q_i''s) and $r' = r$ but $\sum q_i' < \sum q_i$.
We denote $N_s(q, q, ..., q; r)$ by $N_s(q; r)$.

Theorem 10.10 (Schur 1916)**.** *For any positive integer c, there exists the* Schur number $S(c)$ *such that no matter how we colour $[S(c)]$ by c colours, there are $x, y, z \in [S(c)]$ (allow $x = y$) of the same colour and $x + y = z$.*

Proof. Pick $S = N_c(3; 2)$. For any colouring of $[S]$ with c colours, $\phi : [S] \to [c]$, we colour the graph K_S where ab is coloured by $\phi(|b - a|)$. By the definition of $N_c(3; 2)$, there are $i < j < k$ s.t. ij, jk, ik are of the same colour. Take $x = j - i$, $y = k - j$, and $z = k - i$.
(Note: It is better to draw the picture of this colouring of K_S as we did in class.) □

Note: This can be easily extend to a proof where one can find r numbers $x_1, x_2, ..., x_r$ where all the segments $\sum_{i=a}^{b} x_i$ are of the same colour. And the even more general theorem is often attributed to Folkman.

Theorem 10.11. *For any positive integer c and r, there exists $n = n(c, r)$ such that no matter how we colour $[n]$ by c colours, there are $x_1, x_2, ..., x_r \in [n]$, $\sum x_i \leq n$, such that all the $2^r - 1$ sums*

$$\sum_{i \in S} x_i : S \subseteq [r], S \neq \emptyset$$

are of the same colour.

The biggest Schur number we know so far is $S(4) = 45$.
Schur conjectured " a simple lemma, that belongs more to combinatorics than to number theory". Later the same was conjectured by Baudet and proved by van der Waerden when he was 23 years old. Here is the now famous

Theorem 10.12 (van der Waerden 1927)**.** *For any c, l, there is $W = W(c, l)$ s.t. any c-colouring of $[W]$ contains a monochromatic arithmetic progression of length l.*

According to van der Waerden himself, he learned Schur's theorem as late as 1995.
Schur's theorem states that when N is large enough, any c colouring of $[N]$ contains one colour with a solution $x+y-z = 0$. His student Rado generalized Schur's theorem in 1933, here we present one version.

Theorem 10.13 (Rado 1933)**.** *Consider a linear equation $E : \sum a_i x_i = 0$, where all the a_i are integers. Then the following are equivalent:*

(a) For any $c > 0$, there exists $N = N(c)$ such that any c colouring of $[N]$ contains a solution to E where $x_i \in [N]$ and all the x_i's are of the same colour.
(b) There is a non-trivial 0-1 solution to E.

Erdös and Szekeres worked on the similar problem in geometric setting posed by Esther Klein. The paper, published in 1935, includes the result on monotone subsequence we discussed earlier, the Ramsey theorem, and the following

Theorem 10.14 (Erdös-Szekeres 1935)**.** *For any $k > 0$, there is $N = N(k)$ s.t. any N points in the plane in general position (no 3 collinear) has k points that form a convex polygon.*

Four years later Esther Klein became Esther Szekeres. So naturally Erdös called the problem *the happy ending problem.* [3]
Here is a proof due to Tarsy from an exam. The proof was forced out of Tarsy since he did not attend the class when the original proof was taught.

Lemma 10.15. *k points in the plane in general position, they form a convex k-gon iff any 4 of the k points form a convex 4-gon.*

To prove the lemma, use any triangulation of the convex hull.

Proof. (first proof of E-S, due to Tarsy 1972) Pick $N = R(k, k; 3)$, and for N points in general position, colour $\{a, b, c\}$ $(a < b < c)$ yellow if abc is clockwise, otherwise blue. By the definition of N, we have k points whose all triangles are of the same colour. It is easy to see (check it) this implies that any 4 points among the k are in convex position. And therefore, by the lemma, the k points are in convex position. □

Here is a simpler proof that gives a weaker bound. It requires another simple lemma.

Lemma 10.16. *Any 5 points in the plane in general position contains 4 points that form a convex 4-gon.*

[3] During the W.W.II, George and Esther escaped to China and lived in Hong Kou district, Shanghai. George worked in factories. Later they moved to Australia. They passed away on the same day Aug. 28, 2008, within one hour of each other.

Proof. (second proof of E-S) We claim $N = R(k, 5; 4)$ is big enough. Given any N points in the plane in g.p., we colour a 4 point sets as Y if they are convex, B if they are not. Now we have either k points whose all subsets of size 4 are coloured yellow, which implies they are in convex position (Lemma10.15), or 5 points none of its 4-subsets are coloured yellow, which is impossible (Lemma10.16). □

> *I don't feel the least humble before the vastness of the heavens. The stars may be large, but they cannot think or love; and these are qualities which impress me far more than size does.*
>
> – Frank P. Ramsey (1903 - 1930)

11 The basic probabilistic method

The first problem in this section is a cute result from Erdös. [4]

Theorem 11.1 (Erdös 1965)**.** *Let $A = \{a_1, ..., a_n\}$ be a set of n non-zero integers. Prove that there is a subset $B \subseteq A$ such that $|B| > n/3$ and B is sum-free (i.e., no $a, b, c \in B$ with $a + b = c$).*

Proof. Pick a prime number $p = 3k + 2$ such that p is big enough (bigger than all the a_i's). Consider a matrix of n rows and $p - 1$ columns. Put a check on the position (i, x) if $a_i x$ (the element in $\mathbf{Z}_p$) is between $k + 1$ and $2k + 1$, inclusive (*).
The numbers in each row are all different, so, more than 1/3 proportion of each row is checked. So, there exists a column where more than 1/3 of it is checked. i.e., there exists an x such that the set

$$B = \{a_i | k + 1 \leq a_i x \leq 2k + 1\}$$

has more than $n/3$ elements. We claim there are no $a, b, c \in B$ with $a+b = c$. Otherwise, $ax + ay \equiv az \pmod{p}$, but this is impossible by (*). □

The bound was improved slightly by an application of harmonic analysis, but the following is still open:

Open Problem 11.1. *Can we always find such a B with $|B| > n/3+10$ in the above theorem?*

There are certainly many things one may appreaciate in this beautiful proof. Among them, it reveals a bit of probabilistic flavor. One can certainly do without the probability theory, but even with the simple case, the words like "more than 1/3 proportion" let us avoid some trivial counting and divisions. Let us reformulate relevant part of the proof with more probabilistic vocaburary. The probability space consists of simply the numbers from $[p - 1](= [3k + 1])$. Each point has mass $1/(p - 1)$. We pick one x from $[p - 1]$. We define the random variables

$$X_i = \begin{cases} 1 & \text{if } k + 1 \leq a_i x \leq 2k + 1 \\ 0 & \text{otherwise} \end{cases}$$

[4] In 1995, this problem is used in an exam for Chinese mathematics olympiad traning team. During the whole month, among about a hundred exam problems, this is the only one that none of us solved.

We have $\mathsf{E}(X_i) = P(k+1 \leq a_i x \leq 2k+1) > 1/3$. We define the r.v. $X = \sum_{i=1}^{n} X_i$ be the number of a_i's such that $a_i x$ falls into the right interval. By *linearity of expectation*,

$$\mathsf{E}(X) = \sum_i \mathsf{E}(X_i).$$

So, $\mathsf{E}(X) > n/3$, and there is one outcome x where $X > n/3$.

Example 11.1. *Let $G = (V, E)$ be a graph with n vertices and m edges, then we can always partition V into $R + B$ such that there are at least $m/2$ edges between R and B.*

Proof. Randomly colour each vertex as red and blue. (So there are 2^n outcomes in the probability space, each with probability mass $1/2^n$.) For each e, define an indicator r.v.

$$X_e = \begin{cases} 1 & \text{if the endpoints of } e \text{ have different colours} \\ 0 & \text{otherwise} \end{cases}$$

And the r.v. $X = \sum_{e \in E} X_e$ be the number of edges between red and blue. Clearly $\mathsf{E}(X_e) = 1/2$, so $\mathsf{E}(X) = m/2$, there is at least one outcome where $X \geq m/2$. □

Exercise 11.1. *Prove the theorem without the probabilistic method. Consider the partition that maximizes the number of edges between the two parts.*

Exercise 11.2. *Improve the theorem a little bit, by picking a random "almost equal size partition".*

Return to the Ramsey numbers $r(y, b)$ (esp. $r(k, k)$). Now we come to the lower bound. If we can show there is a graph on n vertices and a Y-B colouring of the edges where there is no monochromatic K_k, then $r(k, k) > n$.

Theorem 11.2 (Erdös 1947)**.**

$$r(k.k) \geq \sqrt{2}^k.$$

Proof. Randomly colour each of the $\binom{n}{2}$ edges of K_n Y or B with equal probability. There are $\binom{n}{k}$ possible K_k's, for each of them, the probability it

is monochromatic is $2^{-\binom{k}{2}+1}$, so the probability that at least one of them is monochromatic is at most

$$\binom{n}{k}2^{-\binom{k}{2}+1} < 1$$

It is easy to see (do it) that this is always true when $n < 2^{k/2}$ (in fact the probability is much smaller than 1). So, there is a positive probability for the event "none of the K_k's are monochromatic". i.e., there is one colouring of K_n which does not contain a monochromatic K_k. □

This is almost the best known lower bound today for $r(k,k)$, compare to the best known lower bound from constructive methods, which is less than $k^{\log k}$. (Something to think about: Can Erdös' simple proof also be considered as "constructive"?)
To use the linearity of expectation, for every k-subset $S \subset [n]$, we may define the indicator r.v.

$$X_S = \begin{cases} 1 & \text{if the edges among the points in } S \text{ have the same colour} \\ 0 & \text{otherwise} \end{cases}$$

And $X = \sum_S X_S$. What we proved is that

$$\mathsf{E}(X) = \sum \mathsf{E}(X_S) = \binom{n}{k}2^{1-\binom{k}{2}} < 1.$$

So there is one outcome where $X < 1$, i.e., $X = 0$, and no K_k are monochromatic.
From what we covered in the lectures, we know that

$$\sqrt{2} \leq \liminf r(k,k)^{1/k} \leq \limsup r(k,k)^{1/k} \leq 4$$

both bounds were never improved in decades,

Open Problem 11.2. *Does* $\lim_{k\to\infty} r(k,k)^{1/k}$ *exist? And if so, find the limit.* [5]

[5]Erdös offered \$100 for the first question, and \$500 for the second.

One can argue that this proof can still be made without probability: Draw the $\binom{n}{k}$ rows (all the possible k-points) and $2^{\binom{n}{2}}$ columns (all the possible colourings), mark a cell if the colouring is bad for the row. We can count the number of marked cells and show that it is simply less than the number of columns, so there must be one colouring which is good for all rows. However, we will soon see solutions where the similar "non-probabilistic" arguments are extremely tedious and artificial, while probability is exactly the right language. Let us start with the following example.
Consider $r(k, 4)$. If we mimic the proof above, what we need for n is (a) $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, which can still give us a good n; *but* (b) $\binom{n}{4}2^{1-\binom{4}{2}} < 1$, which will allow us to get only a small n that does not even grow with k. The problem is that it is much easier to get a K_4 than K_k, a lot of energy is wasted on fighting against the K_k's.

Theorem 11.3. *$r(k, 4) \in \Omega(k^{1.5-\epsilon})$ for any $\epsilon > 0$.*

Proof. (scratch) Colour each edge yellow with probability p (and blue with probability $q = 1-p$). Similar to the proof for the symmetric case, it is good enough if

$$\binom{n}{k}p^{\binom{k}{2}} + \binom{n}{4}q^6 < 1.$$

We try to bound both parts by $1/2$. For the second part, it is good enough to take $q = cn^{-2/3}$ for some constant c. Then for the first part,

$$\binom{n}{k}(1 - cn^{-2/3})^{\binom{k}{2}} < \binom{n}{k}e^{-cn^{-2/3}\binom{k}{2}} < (?)1$$

It is then easy to show (fill in the details) $n = \Omega(k^{1.5-\epsilon})$ is good enough for any $\epsilon > 0$. □

The best known lower bound to this day is

$$r(k, 4) \in \Omega(k^{2.5}/\log^2 k).$$

The order for $r(k, 3)$ was open for 60 years and completely solved as

$$r(k, 3) \in \Theta(k^{1.5}/\log k)$$

based on pobabilistic method with 30 pages of calculation (Kim 1995). A "simpler" proof was found later (Bohman 2008), but it is still open if one can find a proof with less than 10 pages.

In many situations the probabilistic method involves some calculations. It takes exercise, experience, courage, and often luck to get good yet elegant estimate on the bounds. We present some usual tools.

Fact 11.4 (the Stirling formula)**.**

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \le n! \le e^{\frac{1}{12n}}\sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Fact 11.5.

$$(1-x)^k \le e^{-kx}$$

We may improve the lower bounds slightly by the *alteration method*. Follow the proof of Theorem 11.2, we have, for any integer n,

$$\binom{n}{k}2^{1-\binom{k}{2}} < 1 \Rightarrow r(k,k) > n. \tag{7}$$

But the proof tells more: for any n, if we do the random colouring, let X as before be the number of monochromatic K_k's, then $\mathsf{E}(X) = \binom{n}{k}2^{1-\binom{k}{2}} =: t$. This tells us that there exists one colouring where there are at most t monochromatic K_k's. Now grab this colourng and say good-bye to the probability space. Remove one point from each of the t subsets, and we get a colouring of the complete graph on $n-t$ points where there is no monochromatic K_k. So,

Theorem 11.6. *For any integer n,*

$$r(k,k) > n - \left\lfloor \binom{n}{k}2^{1-\binom{k}{2}} \right\rfloor.$$

It is easy to see this is stronger than (7). (It turns out this gives a $\sqrt{2}$ times improvement to the lower bound for $r(k,k)$ we can get without alteration.) Similarly, for the general Ransey numbers,

Theorem 11.7. *For any integer n and any $0 \le p \le 1$,*

$$r(k,l) > n - \left\lfloor \binom{n}{k}p^{\binom{k}{2}} \right\rfloor - \left\lfloor \binom{n}{l}(1-p)^{\binom{l}{2}} \right\rfloor.$$

So far we used the fact that, if X is a r.v. in a probablity space, there is always some point where $X \le \mathsf{E}(X)$, the average of X. More generally, here is one of the basic tools,

Fact 11.8 (Markov's inequality)**.** *Let X be a r.v. in a probability space such that $X \geq 0$ everywhere, and let $t > 0$ be a real number, then*

$$\mathsf{Pr}(X \geq t) \leq \mathsf{E}(X)/t.$$

Proof. Prove by contradiction. If $\mathsf{Pr}(X \geq t) > \mathsf{E}(X)/t$, then $\mathsf{E}(X) \geq t\mathsf{Pr}(X \geq t) + 0\mathsf{Pr}(X < t) > \mathsf{E}(X)$. □

For the next problem, let us define (review) some more terms in graphs.

Definition 11.1. *The* girth *of a graph is the length of the shortest cycle in the graph, or ∞ if there is no cycle in the graph.*

For example, K_n has girth 3 for $n \geq 3$; the cude Q_n ($n \geq 2$) has girth 4; the Petersen graph has girth 5.

Definition 11.2. *Let H be a graph. A graph G is called H-*free *if H is not isomorphic to any subgraph of G.*

For example, the Petersen graph is triangle-free. Slightly expand the notation, the bipartite graphs are exactly those odd-cycle-free graphs.

Definition 11.3. *Let $G = (V, E)$ be a graph, $S \subseteq V$ is called a* clique *if $xy \in E$ for any $x, y \in S$. The* clique number *$\omega(G)$ is defined as the size of the largest clique in G.*

Definition 11.4. *Let $G = (V, E)$ be a graph, $S \subseteq V$ is called an* independent set *(or a* stable set*) if $xy \notin E$ for any $x, y \in S$. The* independence number *(or the* stability number*) $\alpha(G)$ is defined as the size of the largest independent set in G.*

Clearly, any clique in G is an independent set in $\overline{G}$, vice versa.

Definition 11.5. *$G = (V, E)$ a graph, and C a set of colours. A (vertex) colouring of G by C is a function $s : V \to C$. A colouring is* proper *if $s(u) \neq s(v)$ whenever $uv \in E$. The graph is said to be k-*colurable *if it has a proper colouring by $[k]$. The* chromatic number *$\chi(G)$ is defined to be the smallest k such that G is k-colourable.*

One way to view the vertex colouring by c colours is that V can be partitioned into c subsets, such that no edge can stay inside the same subset. i.e., V can be partitioned into c stable sets. Immediately, we have

Fact 11.9.

$$\chi(G) \geq n/\alpha(G)$$

It is easy to see this bound is not tight.
$\chi(G) = 1$ iff the graph itself is stable, i.e. the empty graph. $\chi(G) = 2$ iff it is bipartite and not empty – this is almost the definition of the bipartite graph.
Another easy lower bound on $\chi(G)$ is that $\chi(G) \geq \omega(G)$. One way to make $\chi(G)$ big is to put a large clique in G. However, the truth is that there are other constructions, there are even triangle-free graphs with large $\chi(G)$. In fact, here is one of the classics in probabilistic method.

Theorem 11.10 (Erdös 1959)**.** *For any positive integer k and l, there is a graph G s.t. girth of G is bigger than l, and $\chi(G) > k$.*

Definition 11.6. *For any positive integer n and any $0 \leq p \leq 1$, we define the probability space of* random graphs $\mathcal{G}_{n,p} = (\Omega, P)$*, where Ω has $2^{\binom{n}{2}}$ outcomes – all the possible graphs on $[n]$. And for any graph G on $[n]$ with m edges, the probability mass is*

$$P(G) = p^m(1-p)^{\binom{n}{2}-m}.$$

Clearly, this probability space can be specified as: for any $i < j$, randomly pick ij as one edge in G with probability p.
We first try to sketch the proof of the theorem. We will prove that there is a graph on $[n]$ with girth $> l$ and $\alpha(G) < n/k$. Consider random graphs $\mathcal{G}_{n,p}$, where n and p TBD. n^i is an over-estimate of the number of possible cycles in K_n (The actual number is $(n)_i/2i$). Define the r.v. X to be the number of cycles of length at most l, then

$$\mathsf{E}(X) < np + (np)^2 + \cdots + (np)^l.$$

This is a happy news if we pick p to be something like $\frac{1}{2n}$.
Now consider the independent sets of size $t := n/k$. There are $\binom{n}{t}$ such candidates, so

$$\mathsf{Pr}(\exists \text{ stable set of size } t) \leq \binom{n}{t}(1-p)^{\binom{t}{2}} < n^t e^{-pt(t-1)/2} = (ne^{-p(t-1)/2})^t.$$

Plug in $p = 1/2n$, the r.h.s. is something grows fast. However, it is easy to see that we can get a good bound as soon as $p = c\log n/n$ for some reasonable constant c. In this case, $\mathsf{E}(X)$ grows, but not too fast, which means there are not so many short cycles in the graph, and we can do alteration.

Proof. (of Theorem 11.10) For any n, pick $p = 10k \log n/n$, consider the random graphs $\mathcal{G}_{n,p}$. Let the r.v. X be the number of cycles of length at most l, then

$$\mathsf{E}(X) \leq np + (np)^2 + ... + (np)^l < l(10k)^l \log^l n \in o(n).$$

So, when n is big enough, $\mathsf{E}(X) < n/4$, and by Markov,

$$\mathsf{Pr}(X \geq n/2) < 1/2. \tag{8}$$

On the otherhand, let $t = n/2k$. (Well, we can always pick $2k|n$.)

$$\mathsf{Pr}(\exists \text{ stable set of size } t) < (ne^{-p(t-1)/2})^t < 1/2,$$

when n is big enough. So, $\mathsf{Pr}(X \geq n/2 \vee \alpha \geq t) < 1$, so

$$\mathsf{Pr}(X < n/2 \wedge \alpha < t) > 0,$$

which means there is a graph of order n where there are at most $n/2$ short cycles and no independent set of size bigger than t. We delete one vertex from each short cycle, get another graph G on at least $n/2$ vertices where there is no cycle of length less than l, and $\chi(G) > \frac{n}{2}/\frac{n}{2k} = k$. □

We may use the linearity of expectation to simplify the proof: Let X be the number of short cycles in $G \in \mathcal{G}_{n,p}$ and Y be the number of stable sets of size t, with the similar caculation,

$$\mathsf{E}(X + Y) \leq \left(np + (np)^2 + ... + (np)^l\right) + \left((ne^{-p(t-1)/2})^t\right) < n/2$$

when n is big enough. So there exists such a graph with at most $n/2$ short cycles and independent t-sets in total. We then remove one vertex from each short cycle or independent t-set.
We mention that there are constructions of graphs with large girth and large chromatic number, but it is not easy. We leave it as an exercise to construct a triangle-free graph with arbitary large chromatic number.

Definition 11.7. *For a hypergraph $\mathcal{H} = (V, E)$ and a set C of colours, a* proper (vertex) colouring *is a function $s : V \to C$ such that there is no monochromatic hyperedge. A hypergraph is called* k-colourable *if it has a proper colouring by $[k]$. A hypergraph is said to have* property B *if it is* 2*-colourable.*

For graphs (2-uniform hypergraphs), those having property B are exactly the bipartite graphs. We are mainly interested in the property B for r-uniform graphs.

Definition 11.8. *For each r, define $m(r)$ to be the minimum m such that there is a r-uniform hypergraph that does not have property B.*

It is easy to see $m(r)$ is well defined: Consider the complete hypergraph $\binom{[2r-1]}{r}$, so $m(r) \leq \binom{2r-1}{r}$. Clearly $m(2) = 3$. For 3-uniform graphs, $m(3) = 7$. The upper bound is provided in the following example,

Example 11.2 (The Fano configuration)**.** *Figure 6 depicts the finite projective plane of order 2, a.k.a. the Fano configuration. The 3-uniform hypergraph has 7 points and 7 hyperedges.*

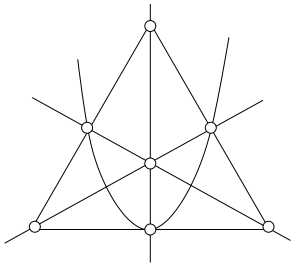

Figure 6: The Fano configuration.

It is easy to check that the Fano configuration does not have property B.
In order to prove that $m(r) \geq m$, we need to show that for any r-uniform hypergraph with less than m edges, we can find a 2-colouring such that there is no monochromatic hyperedge. Let us just randomly colour each vertex, and for each hyperedge e, define X_e to be the indicator r.v. for the event that e is monochromatic. And the r.v. $X = \sum X_e$, the number of monochromatic edges in the outcome. Then

$$\mathsf{E}(X) = m2^{1-r}.$$

There exists an outcome where $X = 0$ when $m < 2^{r-1}$. So,

Theorem 11.11 (Erdös 1963)**.** $m(r) \geq 2^{r-1}$

It is easy to see the equality does not hold.
For the upper bound, we want to find m set of size r each, such that the hypergraph is not 2-colourable. Consider n points (n TBD) and randomly choose m subsets A_1, A_2, ..., A_m of size r. For any r-subset S and any i, $\mathsf{Pr}(A_i = S) = 1/\binom{n}{r}$. Let $\mathcal{H}$ be the hypergraph $([n], \{A_1, A_2, ..., A_m\})$.
There are 2^n vertex 2-colourings. For each colouring s, define the indicator r.v. X_s for the event "s is a proper 2-colouring for $\mathcal{H}$". Suppose s coloured y points yellow and $b = n - y$ points blue, [6]

$$\mathsf{E}(X_s) = \left(1 - \frac{\binom{y}{r} + \binom{b}{r}}{\binom{n}{r}}\right)^m < \left(1 - \frac{\binom{\lfloor n/2 \rfloor}{r}}{\binom{n}{r}}\right)^m < e^{-pm},$$

where $p = \binom{\lfloor n/2 \rfloor}{r}/\binom{n}{r}$. Define $X = \sum X_s$ to be the number of proper 2-colourings for $\mathcal{H}$, so $\mathsf{E}(X) < 2^n e^{-pm}$. We are done if $\mathsf{E}(X) < 1$, i.e. $n \log 2 - pm < 0$. This allows us to make m as small as $n \log 2/p$. The rest of the job is to optimize n so it will give us m as small as possible (in terms of r).

$$\frac{n \log 2}{p} = \frac{n \log 2 (n)_r}{(\lfloor n/2 \rfloor)_r} < n \log 2 \left(\frac{n - r + 1}{n/2 - r + 1/2}\right)^r < 2n2^r(1 + \frac{r}{n - 2r + 1})^r.$$

Note that $(1+1/r)^r \to e$, it is easy to see the last quantity is optimized when n is in the order of r^2, and we have $m(r) \in O(r^2 2^r)$. Just a little bit more work in carefully writing down the proof, one can get

Theorem 11.12 (Erdös 1964)**.** $m(r) < (1 + o(1))e(\log 2)r^2 2^{r-2}$.

For the specific values, we only know $m(r)$ for $r \leq 3$. $m(4)$ is between 20 and 23. For the asymptotics,

Conjecture 11.1 (Erdös - Lovász)**.** $m(r) = \Theta(r2^r)$.

J. Beck improved $m(r)$ to $\Omega(r^{1/3}2^r)$ in 1978 using alterations, based on his proof, the lower bound was improved in 2000 with some more tricks.

Theorem 11.13 (Radhakrishnan - Srinivasan 2000)**.**

$$m(r) \in \Omega(2^r(r/\log r)^{1/2}).$$

[6] My preferrence of the yellow colour is explicit in this notes. I apologize for any confusion it may raise, since people usually pick red and blue as two colours. However, in this proof, yellow is indeed better than red, lest we have an embarrasing $\binom{r}{r}$ in the proof.

Proof. Let p TBD, we consider the probability space as follows. It is the product of 3 spaces. In the first one, for each vertex we randomly independently flip a fair coin, a_v, $\mathsf{Pr}(a_v = R) = \mathsf{Pr}(a_v = B) = 1/2$. In the second one, for each vertex v we randomly independently flip a biased coin with $\mathsf{Pr}(b_v = Y) = p$ and $\mathsf{Pr}(b_v = N) = 1 - p$. And in the last one, we uniformly randomly pick a permutation of all the vertices, $\sigma \in S_n$.
Throughout the proof, remember that the probability space has $4^n n!$ points (possible outcomes), and in probabilitic language, an *event* is just a set of these points.
Now, under each outcome $(\{a_i\}, \{b_i\}, \sigma)$, we produce a colouring by the following algorithm.

```
1.  Colour each vertex v with R or B, based on a_v.
2.  F := monochromatic edges after step 1.
3.  for i := 1 to n
4.     v := σ(i)
5.     if ∃f ∈ F, s.t.  v ∈ f and b_v = Y
6.        switch the colour of v.
7.        for any v ∈ f', F := F − {f'}.
```

We are happy if we can bound $\mathsf{Pr}(\text{there is a red edge at the end}) < 1/2$. For each edge e, define the events

$$A_e = e \text{ is red after step 1, and stays red at the end}$$

$$C_e = e \text{ has blue points after step 1, but all of them changed to red}$$

Note that each vertex gets at most one chance to switch the colour. It is easy to see that, if e is red at the end, then one of A_e and C_e happens.
$\mathsf{Pr}(A_e) = 2^{-r}(1-p)^r$, because A_e happens iff for all $v \in e$, $a_v = R$ and $b_v = N$. (You may look at the algorithm to see why this is "iff".)
Now, for any hyperedge f, define $B_{e,f} = \emptyset$ if $|e \cap f| \neq 1$. Otherwise, let v be the common point of e and f, the event $B_{e,f}$ (e blames f) is defined to be the intersection of the following events
(E_1) $\forall x \in f$, $a_x = B$. (f is blue after Step 1.)
(E_2) $\forall x \in f$ and $x \prec_\sigma v$, $b_x = N$. (f is still in F when we reach v in Step 4.)
(E_3) $b_v = Y$. (v changed from blue (by (i)) to red.)
(E_4) $\forall x \in e$ and $v \prec_\sigma x$, $a_x = R$.
(E_5) $\forall x \in e$ and $x \prec_\sigma v$, either $a_x = R$, or $b_x = Y$.
where $x \prec_\sigma y$ means x comes before y in σ.

We claim that in any outcome where C_e happens, $B_{e,f}$ happens for some f as well, so the event (remember events are just sets of outcomes) $C_e \subseteq \cup_f B_{e,f}$ [7], therefore $\mathsf{Pr}(C_e) \leq \sum_f \mathsf{Pr}(B_{e,f})$.
Consider any outcome where C_e happens, there are blue points in e after Step 1. Let v be the last blue one according to σ, and let f be the one in Step 5 when we loop to v. Because e is finally red, (E_3) holds. Because v was blue, and f was added to F in Step 2, (E_1) holds. Because f is still in F when we reach v, so we have (E_2). (E_4) because v is the last blue point from e according to σ. (E_5) because e is finally red.
We still need to prove that v is the only common point of e and f. Suppose there is another $v' \in e \cap f$, $v' \in f$ implies it is blue after Step 1 (by E_1); then v' comes before v in σ; When we loop to v', because e is finally red, v' changed its colour in Step 6 and f gets removed from F in Step 7, before we loop to v. A contradiction.
Now we bound $\mathsf{Pr}(B_{e,f})$. Condition on any fixed σ, define the r.v. $i = i(\sigma)$ to be the number of points in e that are before v, and j to be the number of points in f that are before v. Note that (E_1) to (E_5) are mutually independent when conditioned on any σ. (Check this! Here you need the fact that v is the only common point of e and f.) So,

$$\mathsf{Pr}(B_{e,f}|\sigma) = \mathsf{Pr}(\bigcap_{1\leq t\leq 5} E_t|\sigma) = \prod_{1\leq t\leq 5} \mathsf{Pr}(E_t|\sigma).$$

We get

$$\mathsf{Pr}(B_{e,f}|\sigma) = \left[2^{-r}\right]\left[(1-p)^j\right][p]\left[2^{-(r-1-i)}\right]\left[(\frac{1}{2}+\frac{p}{2})^i\right]$$

The r.h.s. has 5 parts, represent the conditional probability of the E_i's. So,

$$\mathsf{Pr}(B_{e,f}) = \sum_\sigma \mathsf{Pr}(\sigma)\mathsf{Pr}(B_{e,f}|\sigma) \leq 2^{-2r+1} p \mathsf{E}_\sigma[(1+p)^i(1-p)^j] \leq 2^{-2r+1}p.$$

the last step is itself a cute problem and is left as homework. The rest of the solution is computations to find the optimal p.

[7] It is not necessary for the equality to hold here. The main reason is that in E_5, in order for the outcome to be in C_e, for each x where $a_x = B$, we also need x to be in some other all-blue edge in addition to $b_x = Y$.

$$\begin{aligned}
\mathsf{Pr}(\exists\text{red edge at the end}) &= \mathsf{Pr}(\bigcup_e A_e \cup \bigcup_e C_e) \\
&\leq \sum_e \mathsf{Pr}(A_e) + \sum_e \mathsf{Pr}(C_e) \\
&\leq \sum_e \mathsf{Pr}(A_e) + \sum_{e,f} \mathsf{Pr}(B_{e,f}) \\
&\leq m2^{-r}(1-p)^r + m^2 2^{-2r+1} p <?1/2
\end{aligned}$$

The last inequality holds when $ke^{-pr} + k^2 p < 1$, where $k = m2^{1-r}$. Fix k, it is minimized at $p = \frac{\log(r/k)}{r}$, and we are done if $k^2(1 + \log(r/k))/r < 1$. This implies k can be as big as order $\sqrt{r/\log r}$. □

12 Lovász local lemma

Let A_1, A_2, ..., A_n be bad events, with each $\mathsf{Pr}(A_i) \leq p < 1$. Let the event $B = \cup_i A_i$ be the event that at least one of the bad events happens. Without any other information, we can bound $\mathsf{Pr}(B) \leq np$. So we can say when $n < 1/p$, there exists an outcome where no bad events happens. On the other hand, if we know that all the A_i's are mutually independent, then

$$\mathsf{Pr}(\overline{B}) = \mathsf{Pr}(\bigcap_i \overline{A_i}) \geq \prod_i \mathsf{Pr}(\overline{A_i}) = (1-p)^n$$

So n can be arbitary large, not depend on p. There are situations where the bad events are "almost independent", where they are only dependent "locally".
Let us formally review some vocaburaries on independent events.

Definition 12.1. *Two events A and B are said to be* independent *if*

$$\mathsf{Pr}(A \cap B) = \mathsf{Pr}(A)\mathsf{Pr}(B).$$

If A and B both have positive probability, note that $\mathsf{Pr}(A|B) = \mathsf{Pr}(A \cap B)/\mathsf{Pr}(B)$, and $\mathsf{Pr}(B|A) = \mathsf{Pr}(A \cap B)/\mathsf{Pr}(A)$, we get the following form of independence

$$\mathsf{Pr}(A|B) = \mathsf{Pr}(A), \quad \mathsf{Pr}(B|A) = \mathsf{Pr}(B).$$

These forms makes the term "independent" clearer – knowing B happens does not change the probability of A happens. In fact we can say more – knowing B happens or *not* does not change the probability of A:

$$\mathsf{Pr}(A|\overline{B}) = \mathsf{Pr}(A), \quad \text{and} \quad \mathsf{Pr}(A \cap \overline{B}) = \mathsf{Pr}(A)\mathsf{Pr}(\overline{B}).$$

We omit the formal proof, and being satisfied by drawing shaded areas on the continental map. Certainly, we also have $\mathsf{Pr}(\overline{A}|\overline{B}) = \mathsf{Pr}(\overline{A})$, etc. (Just be careful that, all these are meaningful only when they are well defined, i.e., no division by 0 happens.)

Example 12.1. *Throw a dice 10 times, let A be the event that the 3rd time we get a number less than 3, and B be the event that the sum of the 5th and 6th number is 4. Then $\mathsf{Pr}(A) = 1/3 (= 2 \cdot 6^9/6^{10})$, $\mathsf{Pr}(B) = 1/12$, and $\mathsf{Pr}(A \cap B) = 1/36 (= 2 \cdot 3 \cdot 6^7/6^{10})$. A and B are independent.*

In the previous example, one may say the events are clearly independent because they are defind to be so. The probability space is the product of several independent worlds, the events A and B are sitting in different worlds. It is not so in the following examples.

Example 12.2. *In a bridge hand, 52 cards are randomly distributed to 4 players, let A be the event that player 1 gets at least 2 K's; and B be the event that player 2 gets at least 3 K's.*

Here both $\mathsf{Pr}(A)$ and $\mathsf{Pr}(B)$ are positive, and $\mathsf{Pr}(A \cap B) = 0$. So they are not independent. This conforms to the intuition very well, even we change it slightly to the following

Example 12.3. *In a bridge hand, 52 cards are randomly distributed to 4 players, let A be the event that player 1 gets at least 2 K's; and B be the event that player 2 gets at least 1 K.*

The intuition is that player 1 getting any K will hurt the chance of player 2 getting K's. Anyone with doubts please prove this rigorously.

Example 12.4. *Throw a dice twice, A be the event that the first time we get a 3, B be the event that the sum of the two results is 7.*

I think this is not intuitive. A and B are independent by calculating the actual probabilities.

Example 12.5. *Consider C_3 with vertices a, b, and c. Randomly uniformly colour each vertex as Y or B. Let A be the event that a and b have the same colour; B be the event that b and c have the same colour, and C be the event that c and a have the same colour. Any two of the three events are independent.*

Definition 12.2. *A collection of events A_i $1 \leq i \leq n$ is said to be mutually independent if for any subset $S \subseteq [n]$,*

$$\mathsf{Pr}(\bigcap_{i \in S} A_i) = \prod_{i \in S} \mathsf{Pr}(A_i).$$

Note that, in the previous example, the events A, B, and C are pairwise independent, but not mutually independent.

Definition 12.3. *An event A is said to be* independent of a collection of events $\{A_1, A_2, ..., A_d\}$ *if for any $S \subseteq [d]$, A and $\cap_{i \in S} A_i$ are independent, i.e.*

$$\mathsf{Pr}(A \cap \bigcap_{i \in S} A_i) = \mathsf{Pr}(A)\mathsf{Pr}(\bigcap_{i \in S} A_i).$$

Note that, for a collection of mutually independent events, any one of them is independent to the rest of them. However, the above definition does not require any independence among $A_1, ..., A_d$.
To make the definition look nicer, we show that A is actually independent of any combination of whether each A_i happens or not

Fact 12.1. *If A is independent of $\{A_1, ..., A_d\}$, then for any subset $S \subseteq [d]$, and any $B_i = A_i$ or $\overline{A_i}$, A is independent of $\cap_{i \in S} B_i$.*

One can prove this by induction on the number of complemented A_i's. Again, we omit the formal proof.

Definition 12.4. *Let A_1, A_2, ..., A_n be events in a probability space, a digraph G on the A_i's is called* a dependency graph for the A_i's *if for any i, A_i is independent of the collection of events $\{A_j : A_i A_j \notin G, i \neq j\}$.*

Be careful about the formal definition. Strictly, we make a distinct symbol for each event, so the graph has n vertices. Two events might be the same set, but they corresponding to different vertices. When we talk about an event as a vertex in the dependency graph, we are actually talking about its corresponding symbol.
Note that the dependency graph may not be unique. If all the events are mutually dependent, then any graph can be a dependency graph. On the other extremal case, if the events are pairwise dependent, then the only dependency graph is the complete digraph. If G is a dependency graph, then adding edges we still get a dependency graph. So it is harder to get a small dependency graph. The existence of smaller dependency graphs imply the events are more "independent".
The following lemma appeared in a paper by Erdös and Lovász. Erdös called it *Lovász local lemma*.

Lemma 12.2 (Erdös - Lovász 1975)**.** *Let A_1, A_2, ..., A_n be events, and G be a dependency graph. If for each i, $\mathsf{Pr}(A_i) \leq p < 1$ and $d^+(A_i) \leq d$, and $4pd \leq 1$, then*

$$\mathsf{Pr}(\bigcap_{i \in [n]} \overline{A_i}) > 0.$$

Consider the A_i's as "bad events". The lemma says roughly, if the events are almost independent except in a small neighbourhood, then there is at least one outcome where none of the bad events happens. Note that here n can be arbitary big, not depending on p. [8]

Proof.

$$\mathsf{Pr}(\bigcap_{i=1}^{n} \overline{A_i}) = \mathsf{Pr}(\overline{A_1})\mathsf{Pr}(\overline{A_2}|\overline{A_1})\mathsf{Pr}(\overline{A_3}|\overline{A_1} \cap \overline{A_2})...\mathsf{Pr}(\overline{A_n}|\bigcap_{i=1}^{n-1} \overline{A_i})$$

Be careful we want the above formula to be well defined. We actually prove by induction that any of the terms above is positive, so all the conditional probabilities are well defined. Below is the formal proof.
The theorem is trivial when $d = 0$ – the events are mutually independent. When $d > 0$, we show that if we pick any $s + 1$ events A and B_1, B_2, ..., B_s among the n original events, then

$$\mathsf{Pr}(A|\bigcap_{i=1}^{s} \overline{B_i}) \leq 1/2d \quad (*)$$

If we can prove (*), then all the above conditional probabilities are well defined, and the product is at least $(1 - 1/2d)^n > 0$. We prove (*) by induction on s.
Basis. $s = 0$, $\mathsf{Pr}(A) \leq p \leq 1/4d$.
Inductive Step. Suppose $s > 0$ and (*) holds for any $s' < s$. Rename the B_i's to be D_1, D_2, ..., D_x, E_1, E_2, ..., E_y, where $x + y = s$, and the D_i's are all the events where $A \to D_i$ in the dependency graph, so $x \leq d$.
If $x = 0$, by the definition of the dependency graph, $\mathsf{Pr}(A|\bigcap_{i=1}^{s} \overline{B_i}) = \mathsf{Pr}(A) \leq p < 1/2d$, we are done. In the following we assume $x > 0$ so $y < s$.
We have

$$\mathsf{Pr}(A|\bigcap_{i=1}^{s} \overline{B_i}) = \frac{\mathsf{Pr}(A \cap \bigcap_{i=1}^{s} \overline{B_i})}{\mathsf{Pr}(\bigcap_{i=1}^{s} \overline{B_i})} \leq \frac{\mathsf{Pr}(A \cap \bigcap \overline{E_i})}{\mathsf{Pr}(\bigcap \overline{D_i} \cap \bigcap \overline{E_i})} = \frac{\mathsf{Pr}(A|\bigcap \overline{E_i})}{\mathsf{Pr}(\bigcap \overline{D_j}|\bigcap \overline{E_i})} =: q$$

[8] We mention that there are other variations of the local lemma. With a little efforts, with $1/2d$ replaced by $1/(d + 1)$ in the proof, one can replace the condition $4pd \leq 1$ by $epd \leq 1$ in the lemma.

Note that the conditional probabilities in the last step is well defined ($\mathsf{Pr}(\bigcap \overline{E_i}) > 0$) because of the inductive hypothesis.
Now, by the definition of the dependence graph, A is independent of the E_i's, so the numerator is exactly $\mathsf{Pr}(A) \leq p$. By the inductive hypothesis (remember that now $y < s$), $\mathsf{Pr}(B_j | \bigcap \overline{E_i}) \leq 1/2d$, so $\mathsf{Pr}(\bigcup B_j | \bigcap \overline{E_i}) \leq d\frac{1}{2d} \leq 1/2$. Taking the complement, so the denominator of q is at least $1/2$. Therefore $q \leq 2p \leq 1/2d$.

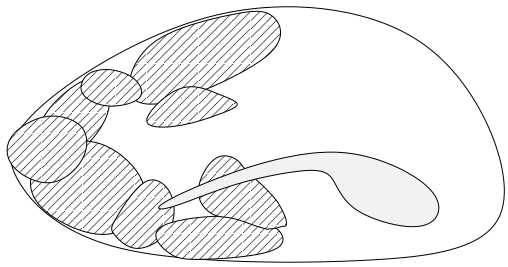

Figure 7: The proof of L.L.L: The world is conditioned on $\bigcap \overline{E_i}$; the gray area is p; there are at most d shaded parts, each has area (by induction) at most $1/2d$, so the gray portion in the unshaded world is at most $p/(1/2) \leq 1/2d$.

□

Example 12.6. *$r(k,k) > n$, if*

$$4\binom{k}{2}\binom{n}{k-2}2^{1-\binom{k}{2}} \leq 1.$$

Solution. Randomly colour the edges of K_n with Y or B. There are $N = \binom{n}{k}$ possible candidates for monochromatic K_k. For each $|S| = k$, let A_S be the event that the edges in S is monochromatic. A_s is independent of the collection $\{A_T : |T \cap S| \leq 1\}$. □
Compare this with the bound we get from the linearity of expecation, where n needs to be

$$\binom{n}{k}2^{1-\binom{k}{2}} < 1.$$

L.L.L. gives a slightly better bound.

Example 12.7. *Suppose $\mathcal{H}$ is a hypergraph where each edge has r elements and meets at most d other edges. If $d \leq 2^{r-3}$, then $\mathcal{H}$ has property B.*

Solution. Randomly colour each vertex red or blue. For each edge e, let A_e be the events that e is coloured monochromatic. For the dependency graph, $A_e \to A_f$ if e intersects f. □

Corollary 12.3. *For any $k \geq 10$, any k-uniform k regular hypergraph has property B.*

Note, we can replace 10 with 9 if we use the *epd* ≤ 1 version of the local lemma.
Consider the algorithmic problem: Given a digraph, decide if there is any cycle of even length in the graph. The problem might seems easy at first glance. It is equivalent to several other hard problems, and opened for decades until around 2000 when people solved it with long proofs.
However, using the local lemma, we have an elegant solution for graphs with enough out-degrees and not too big in-degrees.

Theorem 12.4 (Alon - Linial 1989)**.** *Any k-diregular digraph G with $k \geq 8$ has a directed cycle of even length.*

And more generally,

Theorem 12.5 (Alon - Linial 1989)**.** *Let G be a digraph where $\delta^+ > 0$ and*

$$4\delta^+\Delta^-(1-1/k)^{\delta^+} \leq 1,$$

then G contains a directed cycle whose length is a multiple of k.

Proof. We may assume each out-degree is exactly δ^+ – removing edges and we can get such a graph without increasing Δ^-, and it is enough to find such a cycle in the smaller graph.
If we can partition (colour) the vertices of G into k classes V_0, V_1, ..., V_{k-1}, such that each vertex in V_i has at least one edge going into the next class V_{i+1} (where the addition is done on $\mathbf{Z}_k$, e.g., $(K-1)+1=0$), then we are happy: Start from anywhere, and keep going to the next class, the process must end when the same vertex is repeated for the first time, where we have a simple cycle of length multiple of k.
We just do the random partition. For each vertex v, the bad event A_v is "none of u where $v \to u$ goes to the class next to the class of v". It is clear $\mathsf{Pr}(A_v) = (1-1/k)^{\delta^+}$. Now we just need to find a dependency graph on the A_v's where each out-degree is at most $\delta^+\Delta^-$. For this purpose, we draw an

edge from A_v to A_u if either (i) $u \in N^+(v)$ (there are exactly δ^+ such u's), or (ii) $N^+(u) \cap N^+(v) \neq \emptyset$ and $u \neq v$ (there are at most $\delta^+(\Delta^- - 1)$ such u's). This time let us check that this indeed satisfies the definition of a dependency graph. (Did you forget to do this for the previous examples because they seemed "too obvious"?)

The probability space has k^n points. For a vertex v, we may look at this space by first colour all the points in $V - N^+(v)$, thus partition the universe into $k^{n-\delta^+}$ parts, let's call them potatoes. Inside each potato, we randomly colour points in $N^+(v)$, there are k^{δ^+} possible outcomes. As we see, the conditional probability of A_v inside each potato is $(1 - 1/k)^{\delta^+}$ (*).

For any collectinon A_1, A_2, ..., A_t such that $A_i \neq A_v$ and $A_v \not\to A_i$ in the dependency graph. Inside each potato, the colouring of $V - N^+(v)$ is fixed, so (by i and ii) $\bigcap_{i=1}^t A_i$ either happens or not. This means $\bigcap_{i=1}^t A_i$ is a disjoint union of potatoes. By (*), $\mathsf{Pr}(A_v | \bigcap_{i=1}^t A_i) = \mathsf{Pr}(A_v)$.

□

A few comments on the proof. When we described an edge $A_v \to A_w$ in the dependency graph (or $A_e \to A_f$ in the previous example about property-B), very often these events are pairwise independent (even though sometimes their support share a point). It is usually the combined event of several A_w's that affects the chance of A_v. We also note that, unlike the previous examples, the dependency graph here is indeed a digraph – the dependency relation is not symmetric.

13 Combinatorics of sets

We start with probably the most important theorem about finite sets in matching theory. The hard thing is that it is quite difficult to give it a name. Consider an $n \times m$ 0-1 matrix, when can we pick n 1's such that each row has exactly one 1, and they are from different columns? i.e., find an $n \times n$ permutation matrix inside it. Clearly, any k rows combined must have at least k candidates

Theorem 13.1 (König 1931)**.** *An $n \times m$ 0-1 matrix contains an $n \times n$ permutation matrix* iff *for any k rows there are at least k columns having 1's in them.*

Consider a sequence (repetition allowed) of sets A_1, A_2, ..., A_n, when can we pick x_i from A_i such that all the x_i's are distinct? The set of x_i's is called a *system of distinct representitives*. Clearly, if some A_i is empty, then this cannot be done. And if some $A_i = A_j = \{x\}$, we are also in trouble. In general, if there are k of the A_i's whose union has only less than k elements, then this cannot be done.

Theorem 13.2 (Hall 1936)**.** *A collection of sets A_1, A_2, ..., A_n has a system of distinct representitives* iff *for any $T \subseteq [n]$, $|\bigcup_{i \in T} A_i| \geq |T|$.*

There are n girls and m boys, each girl has several boys (A_i) she is willing to marry. We can make n pairs iff for any k girls, the total number of boys they are willing to marry is at least k. Thus the above theorem is commonly called *Hall's marriage theorem.*
In this section, we denote (A, B) a bipartite graph where all the edges are between A and B.

Definition 13.1. *In a graph, a* matching *is a set S of edges such that any vertex belong to at most one edge in S. In a bipartite graph (A, B), a* perfect matching *is a matching where all the vertices are matched; an* perfect matching for A *is a matching where all the vertices in A are matched.*

Draw the picture for the above 2.5 theorems, we see they are exactly the same. The following is the most illustrative form of these equivalent (well, you may say all the true statements are equivalent...) theorems. In a graph, let us denote $N(X)$ all the neighbours of the vertex set X, i.e., $N(X) := \bigcup_{v \in X} N(v)$.

Theorem 13.3. *In a bipartite graph (A, B), there is a perfect matching for A* iff *for any subset $X \subseteq A$, $|N(X)| \geq |X|$.*

Proof. The "only if" is obvious. We prove the "if" direction by induction on $n := |A|$. Let (*) be the condition

$$\forall X \subseteq A, |N(X)| \geq |X|$$

The base case is trivial. Suppose $|A| = n$, the graph satisfies (*), and the statement is true for any $n' < n$.
Case 1. For any non-empty $X \subset A$ ($X \neq A$, $X \neq \emptyset$), $|N(X)| > |X|$. We pick any $u \in A$, by (*) u has some neighbour $u' \in B$. Match u to u', and it is easy to check the remaining graph on $(A - u, B - u')$ satisfies (*) and thus has a matching of size $n - 1$ by induction.
Case 2. There is non-empty $X \subset A$ where $N(X) = X$. Let $Y = N(X) \subseteq B$, $A' = A - X$, and $B' = B - Y$. By induction (since $|X| < n$) there is a perfect matching between X and Y. And it is easy to check (A', B') satisfies (*). (Let's say "draw it" instead of "prove it".) □

Example 13.1. *Any k-regular bipartite graph has a perfect matching, even when the graph is allowed to have multiple edges.*

By the proposition above, we see that any k-regular bipartite graph can be *covered* by k perfect matchings. And this has the same picture as the following theorem due to Birkhoff.

Theorem 13.4. *Let A be an $n \times n$ matrix with nonnegative integer entries. If each row sum and each column sum is k, then A is the sum of k permutation matrices.*

Definition 13.2. *In a bipartite graph (A, B), for any $X \subseteq A$, the* deficiency *of X is defined to be* $\max(0, |X| - |N(X)|)$.

Hall's theorem says that if all X has deficiency 0, then there is a perfect matching for A. The following is an easy corollary (generalization) of Hall's theorem.

Corollary 13.5. *In a bipartite graph (A, B) where $|A| = n$, the size of the maximum matching equals $n - D$, where D is the maximum deficiency.*

Proof. Clearly the maximum matching cannot be bigger than $n - D$, since some X has only $|X| - D$ neighbours. To show that $n - D$ is achievable, consider adding D vertices to B, connect all of them to every vertex of A, show that the new graph satisfies the condition of Hall's theorem. □

Consider $A - X$ and $N(X)$, it is clear (draw it) that all edges have at least one end in them. So they form a *vertex cover*. Now it is easy to see Hall's theorem is equivalent to

Theorem 13.6 (König 1931)**.** *In a bipartite graph, the size of the maximum matching equals the size of the minimum vertex cover.*

And not a surprise at all, they all fall under the magic spell *max-flow-min-cut*. Recall the definition of a poset.

Definition 13.3. *A* partially ordered set *(or* poset*) is a pair* $\mathcal{P} = (S, \preccurlyeq)$ *where* S *is a set and* $\preccurlyeq$ *is a relation on* S *such that*

- $\forall x \in S, x \preccurlyeq x$*. (*$\preccurlyeq$ *is reflexive.)*
- *If* $x \preccurlyeq y$ *and* $y \preccurlyeq z$*, then* $x \preccurlyeq z$*. (*$\preccurlyeq$ *is transitive.)*
- *If* $x \preccurlyeq y$ *and* $y \preccurlyeq x$*, then* $x = y$*. (*$\preccurlyeq$ *is antisymmetric.)*

Definition 13.4. *In a poset* $\mathcal{P}$*, we write* $x \prec y$ *if* $x \prec y$ *and* $x \neq y$*. A* chain *in a poset is* $(x_1, x_2, ..., x_k)$ *such that* $x_i \prec x_{i+1}$ *for all* $i < k$*. An* antichain *in a poset is a set* S *of elements where no pairs are comparable, i.e.,* $\forall x, y \in S$*,* $x \not\prec y$*. A set of chains* covers $\mathcal{P}$ *if their union is* $\mathcal{P}$*.*

For example, all the *maximal* elements of a poset form an antichain. In the n-dimensional cube, when $\preccurlyeq$ means $\subseteq$, each level $\binom{[n]}{r}$ form an antichain.

Theorem 13.7 (Dilworth 1950)**.** *Let* $\mathcal{P}$ *be a poset,* m *be the minimum number of chains that can cover* $\mathcal{P}$*, and* M *be the size of a maximum antichain, then* $m = M$*.*

Proof. It is trivial that $m \geq M$, since each chain can only contain one element from the maximum antichain. We prove M chains is enough to cover $\mathcal{P}$ by induction on P. As usual, the base case is trivial.
Let A be a maximum antichain in $\mathcal{P} = (S, \preccurlyeq)$, define $U := \{x : \exists a \in A, a \prec x\}$, $L := \{x : \exists a \in A, x \prec a\}$. Since $a \prec x \prec b$ implies $a \prec b$, $U \cap L = \emptyset$. And because A is maximum, we have $S = U + A + L$. If both U and L are

not empty, then we use induction on $U + A$ as well as $V + A$ to get two sets of M chains, and we can combine them by the element in A. (Fill in the details.) Now the only trouble rises when one of them, say, U, is empty.
Now, consider C to be a *maximal* chain in $\mathcal{P}$. If, after removing C, there are no antichains of size M, we are done by induction. So assume there is still an antichain A of size M in $\mathcal{P} - C$. For this A, as we discussed above, we may assume that U is empty, otherwise we are done. Let x be the *maximum* element of C, $x \in L$, then there is $a \in A$ such that $x \prec a$. This means $C + a$ is a chain larger than C. □

Use Dilworth theorem on posets with just two levels, we get another easy proof of Hall's theorem. A more interesting fact is that Dilworth theorem also easily follows from Hall's theorem.

Proof. (Hall ⇒ Dilworth) Form a bipartite (A, B) where $|A| = |B| = |S|$, for each element $x \in S$ we have one a_x in A and b_x in B. Connect a_x with b_y if $x \prec y$. By (the matching-cover version of) König theorem, let t be a maximum matching of size t, then there is a vertex cover C of size t. Clearly $t < n$ since we can take all the r.h.s. minus b_x for some minimal $x \in S$.
Now there are at least $n - t$ elements such that neither a_x nor b_x is in C. It is clear these x's form an antichain of size $\geq n - t$ in S. All we need to prove now is that S can indeed be covered by $n - t$ chains.
For this we look at the matching of size t. There are $n - t$ unmatched points in B, start from any of them, follow the matching edges, we get a chain. And these chains cover S. □

How many subsets of $[n]$ can we pick such that none of them is a subset of another? In other words, what is the size of the maximum antichain in the poset $2^{[n]}$? A quick answer is that we can pick all the subsets of the same size k, e.g. the $\binom{n}{k}$ sets $\binom{[n]}{k}$. We know that this is maximized when $k = \binom{n}{\lfloor n/2 \rfloor}$ or $k = \binom{n}{\lceil n/2 \rceil}$.

Theorem 13.8 (Sperner 1928)**.** *If A_1, A_2, ..., A_m are subsets of $[n]$ such that $A_i \not\subseteq A_j$ whenever $i \neq j$, then $m \leq \binom{n}{\lfloor n/2 \rfloor}$.*

Proof. (using Hall) Consider any two adjacent levels in $2^{[n]}$ and the edges between them (A and B has an edge iff one of them contains the other). All the points on one level have the same degree. It is easy to show that Hall's condition holds for the level with fewer sets.

We can start from the middle level, consider it as $\binom{n}{\lfloor n/2 \rfloor}$ chains (each is a singleton yet now). Use these mathcings to extend the chains upwards while all the sets are covered level by level. Do the same from the middle level downwards.
Since $2^{[n]}$ can be covered by $\binom{n}{\lfloor n/2 \rfloor}$ chains, so there can not be an antichain of bigger size. □

Proof. (via symmetric chain partition) In $2^{[n]}$, a chain of sets $A_1 \subseteq A_2 \subseteq ... \subseteq A_t$ is symmetric if $|A_{i+1}| = |A_i| + 1$ for any $1 \leq i < t$, and $|A_1| + |A_t| = n$. It is enough to show that $2^{[n]}$ has a symmetric chain parittion, because then each symmetric chain must hit the middle level exactly once, and we have a chain cover of size $\binom{n}{\lfloor n/2 \rfloor}$.
We use induction to show the existsence of the symmetric chain partition. Suppose we have such a partition for $2^{[n-1]}$, with chains C_1, C_2, ... C_t. Let $D_i = C_i \cup \{n\}$, then the C_i's and D_i's is a chain cover of $2^{[n]}$. They are nearly symmetric, but not exactly. Assume A_i is the smallest set in C_i, we let

$$C_i' = C_i - \{A_i\}, \text{ and } D_i' = D_i \cup \{A_i\}.$$

It is easy to see the C_i''s and D_i''s gives a symmetric chain partition of $2^{[n]}$. (Note that some of the C_i' are empty and disappear from the picture, this is the reason why the number of chains does not always double from $n-1$ to n.) □

Proof. (proportion of shadows) Let $\mathcal{A}$ be a set of $(r+1)$-subsets of $[n]$, i.e., $\mathcal{A} \subseteq \binom{[n]}{r+1}$. Its *shadow* is defined as

$$\partial\mathcal{A} = \{B \in \binom{[n]}{r} : \exists A \in \mathcal{A}, B \subseteq A.\}$$

It is easy to prove (exercise) that the proportion of $\partial\mathcal{A}$ in its level is at least as big as the proportion of $\mathcal{A}$ in level $(r+1)$, i.e.,

$$\frac{\partial\mathcal{A}}{\binom{n}{r}} \geq \frac{\mathcal{A}}{\binom{n}{r+1}}.$$

Now, for any antichain $\mathcal{F}$ in $2^{[n]}$, define α_r to be the number of r-sets in $\mathcal{F}$. Define

$$f(\mathcal{F}) = \sum_r \frac{\alpha_r}{\binom{n}{r}}.$$

Start from any antichain $\mathcal{F}$, we take the sets in the highest level in $\mathcal{F}$, and replace them with the shadow. It is easy to see that the shadow does not include any existing elements in $\mathcal{F}$. Let $\mathcal{F}'$ be the resulting family, it is easy to see that $\mathcal{F}'$ is again an antichain, and $f(\mathcal{F}) \leq f(\mathcal{F}')$.
Keep doing this, we have all in the trivial situation where $f = 1$. So $f(\mathcal{F}) \leq 1$ for any antichain, and the size of the antichain

$$\sum_r \alpha_r \leq \binom{n}{\lfloor n/2 \rfloor}.$$

□

Proof. (Lubell 1966) Consider any maximal chain $C_0 \subseteq C_1 \subseteq ... \subseteq C_n$ where $|C_i| = i$. i.e., start from $\emptyset$, adding one element in each step. There are $n!$ such chains.
For any A_i, suppose it has size k_i, how many chains has A_i as an element? It must occur on the k_i-th position. There are $k_i!(n - k_i)!$ such chains.
$A_i \not\subseteq A_j$ implies that any of the $n!$ chains can contain only one of the A_i's. So

$$m\lfloor n/2 \rfloor! \lceil n/2 \rceil! \leq \sum_{i=1}^{m} k_i!(n - k_i)! \leq n!.$$

□

Proof. (translated into probabilistic language) Randomly uniformly pick a permutation σ of $[n]$. For each $1 \leq i \leq m$, suppose $|A_i| = k_i$, define X_i to be the indicator r.v. for the event that the first k_i elements in σ form the set A_i. Let $X = \sum X_i$. $A_i \not\subseteq A_j$ implies that at most one of X_i's can be 1. So

$$1 \geq \mathsf{E}(X) = \sum_{i=1}^{m} \mathsf{E}(X_i) = \sum_{i=1}^{m} \frac{k_i!(n - k_i)!}{n!} = \sum_{i=1}^{m} \frac{1}{\binom{n}{k_i}} \geq \frac{m}{\binom{n}{\lfloor n/2 \rfloor}}.$$

□

In both proofs, the equality holds if $k_i = \lfloor n/2 \rfloor$ or $\lceil n/2 \rceil$ for all i. When n is even, clearly there is only one situation we can achieve the maximum number of subsets. It is left as an exercise to prove that when n is odd and the A_i's achieve the maximum number, all the A_i's must have the same size.

Definition 13.5. *A family $\mathcal{F}$ of sets is called* intersecting *if for any $A \in \mathcal{F}$, $B \in \mathcal{F}$, we have $A \cap B \neq \emptyset$.*

How many subsets of $[n]$ can we pick such that any two of them intersect? This question is quite trivial. If we have more than 2^{n-1} subsets, then there are two of them complement each other. On the other hand, there are different examples showing that 2^{n-1} can be achieved. e.g., pick all the subsets containing a fixed element 1. Or, as observed by Zhu Hongyu in class, if n is odd, pick all the subsets with size more than $n/2$.
How many subsets of $[n]$, with the same size k, can we pick such that any two of them intersect? The problem is entirely trivial when $2k > n$, where we can pick all of $\binom{[n]}{k}$. When $2k \leq n$, a quick answer would be: Fix an element, say 1, and pick all the k-subsets that include 1. There are $\binom{n-1}{k-1}$ such sets.

Theorem 13.9 (Erdös - Ko - Rado 1938, published 1961)**.** *Let n and k be positive integers where $2k \leq n$. If A_1, A_2, ..., A_m are subsets of $[n]$ such that $|A_i| = k$ for all i and $\{A_i\}$ is an intersecting family, then $m \leq \binom{n-1}{k-1}$.*

Proof. (Katona 1974) Equally divide the unit circle into n arcs, and randomly uniformly give each arc a distinct label from $[n]$. Each of the $n!$ labelings occur with probability $1/n!$.
For each set A_i, define the r.v. X_i to be the indicator of the event that the k arcs labeled by the elements of A_i are consecutive on the circle. It is easy to calculate that

$$\mathsf{E}(X_i) = n\frac{k(k-1)...1}{n(n-1)...(n-k+1)} = \frac{k!}{(n-1)_{k-1}}.$$

Define $X = \sum_{i=1}^{m} X_i$. $\mathsf{E}(X) = \frac{mk!}{(n-1)_{k-1}}$. Now, the semantic meaning of X is the number of subsets whose elements are labeled on consecutive arcs. In any outcome, suppose these sets are B_1, B_2, ..., B_X. Fix B_1, by the intersecting property, any other B_i either enters on the left side (type L) or enters on the right side (type R), but (since $n \geq 2k$) cannot be both. Again, by $n \geq 2k$, if there is a type L ended at the i-th arc, there cannot be another with type R and started at the $(i+1)$-th arc. So, including B_1, there are at most k such sets.
We have

$$\frac{mk!}{(n-1)_{k-1}} = \mathsf{E}(X) \leq \max(X) \leq k.$$

So $m \leq \binom{n-1}{k-1}$. □

Definition 13.6. *A family of sets $\mathcal{F}$ is called a* star *if there is an element x such that $x \in A$ for all $A \in \mathcal{F}$.*

Definition 13.7. *A family of sets $\mathcal{F}$ is called* downward closed *if for any $A \in \mathcal{F}$, and any $B \subseteq A$, we have $B \in \mathcal{F}$.*

Here is an innocent looking, notoriously hard conjecture: If $\mathcal{F}$ is downward closed, then no intersecting family in $\mathcal{F}$ can be larger than its biggest star.

Conjecture 13.1 (Chvátal 1972)**.** *If $\mathcal{F}$ be a downward closed family of sets with $|\mathcal{F}| > 1$, then the maximum size of an intersecting family in $\mathcal{F}$ equals the maximum size of a star in $\mathcal{F}$.*

14 Some problems in discrete geometry

In the section on Ramsey numbers, we mentioned the Happy Ending problem posed by Esther Klein in 1933, and generalized by Paul Erdös and George Szekeres (Theorem 10.14). The theorem states that, for any n, there is a least number $N(n)$, such that any $N(n)$ points in the plane in general position must contain a convex n-gon. In fact, they proved

Theorem 14.1. *[Erdös-Szekeres 1935] For any $n \geq 3$,*

$$2^{n-2} + 1 \leq N(n) \leq \binom{2n-4}{n-2} + 1.$$

Definition 14.1. *Let $p_i = (x_i, y_i)$ $(1 \leq i \leq t)$ be a sequence of points in the plane such that $x_i < x_{i+1}$ for each i. It is called a $t \smile$ if the slope of $\overrightarrow{p_{i+1}p_{i+2}}$ is bigger than the slope of $\overrightarrow{p_ip_{i+1}}$ for all i. It is called a $t \frown$ if the slope of $\overrightarrow{p_{i+1}p_{i+2}}$ is smaller than the slope of $\overrightarrow{p_ip_{i+1}}$ for all i.*

Lemma 14.2. *For any positive integers p and q, any $\binom{p+q}{p} + 1$ points in the plane in general position with distinct x-coordinates contains a $(p+2) \smile$, or $(q+2) \frown$.*

Proof. Prove by induction. The basis are easy to check. For the inductive step. Let $t_1 = \binom{p+q-1}{p-1} + 1$ and $t_2 = \binom{p+q-1}{p} + 1$, so

$$t := \binom{p+q}{p} = t_1 + t_2 - 1.$$

Now, start from these t points, we pick any t_1 points, by inductive hypothesis, either we find a $(q+2) \frown$ (then we are done), or there is a $(p+1) \smile$. Let v be the rightmost point of this $(p+1) \smile$. We remove v from our point set.
We keep doing this as long as there are at least t_1 points, so we can do this t_2 times, and get t_2 such v's. Let v_1, v_2, ..., v_{t_2} be these points, from left to right.
By inductive hypothesis again, among the v_i's we either find a $(p+2) \smile$ (then we are done), or we find a $(q+1) \frown$. Let w be the leftmost point of this $(q+1) \frown$.
Thus we find w which is the rightmost point of P_1, a $(p+1) \smile$; at the same time the leftmost point of P_2, a $(q+1) \frown$. Let x be the second rightmost point in P_1 and y be the second leftmost point in P_2. Depending on whether the slope of $\overrightarrow{xw}$ is bigger than the slope of $\overrightarrow{wy}$, we can extend x to P_2 to get a $(q+2) \frown$, or y to P_1 to get a $(p+2) \smile$. □

Lemma 14.3. *For any positive integers p and q, there exists configuration $\mathcal{C}_{p,q}$ with $\binom{p+q}{p}$ points, no two has the same x-coordinate, such that there are no $(p+2)\smile$ nor $(q+2)\frown$ in the configuration.*

Proof. Induction on $p+q$. The basis are easy to check. For the inductive step, note that no matter how we rescale a configuration along x or y-axis, $\smile$s are still $\smile$s and $\frown$s still $\frown$s. So we may have a $\mathcal{C}_{p-1,q}$ and a $\mathcal{C}_{p,q-1}$, such that both have a bounding box of 1×1, and inside each configuration any two points form a slope of absolute value at most 0.1.
Now put $\mathcal{C}_{p-1,q}$ around the point $(0,0)$ and $\mathcal{C}_{p,q-1}$ around $(2,100)$, so any point in the former to any point in the latter form a slope at least (much bigger) 1. It is easy to check this combined configuration qualifies as $\mathcal{C}_{p,q}$. □

Proof. (of Theorem 14.1) For the upper, since there are finitely many directions, we may pick one direction as the x-axis so that no two points share the same x-coordinate. Then we apply Lemma 14.2, there is not only a convex n-gon, but an $n\smile$ or $n\frown$, something even stronger.
For the lower bound, we need to show that there is a configuration with 2^{n-2} points where there does not exist any convex n-gon. To setup our frame, we draw S, an $(n-1)\smile$ where any two x-coordinates differs by at least 10 and slope between two points is at least 100. As discussed in Lemma 14.3, we may pick a block $\mathcal{C}_{r,n-2-r}$ $(r = 0, ..., n-2)$, resize it into a unit box, and such that any two points in $\mathcal{C}_{r,n-2-r}$ has slope with absolute value at most 0.01.
Now we put $\mathcal{C}_{r,n-2-r}$ around the $r+1$-st point on S. Consider any convex polygon $\mathcal{P}$, it consists of an upper chain which is a $\frown$, and a lower chain which is a $\smile$. If $\mathcal{P}$ touches only one block, it is easy to check it has less than n points. Otherwise let r_1 be the first block it touches, and r_2 be the last one. The lower chain can touch each block in the middle at most once, while the upper chain cannot touch any blcok in the middle. It is easy to check in this case, again, there are less than n points. □

They liked the construction of 2^{n-2} points without a convex k-gon so much that they made and later Erdös offered 500 dollars for the following

Conjecture 14.1 (Erdös-Szekeres 1935)**.** *For any $n \geq 3$, $N(n) = 2^{n-2} + 1$.*

With the faith in the lower bound, one might expect to improve the seemingly loose upper bound significantly. However, the first improvement came 62 years later. The method is completely different, but the result differs only by 1.

Theorem 14.4 (Chung-Graham 1997)**.** *For $n \geq 3$, $N(n) \leq \binom{2n-4}{n-2}$.*

Further small improvements were made. The best we know today is

Theorem 14.5 (Tóth-Valtr 2005)**.** *For $n \geq 3$, $N(n) \leq \binom{2n-5}{n-2} + 1$.*

How about empty (one that does not contain in its interior other points from the set) convex k-gons? It is easy to show that if the number of points is large enough, one can always find an empty 5-gon. (Thus an empty (≤ 5)-gon.) On the other hand,

Theorem 14.6 (Horton 1983)**.** *For any N, there is a set S of N points in general position in the plane such that any convex 7-gon formed by points in S must also contain some $p \in S$ in its interior.*

So, there was a gap for $k = 6$. This was the long standing empty hexagon problem.

Theorem 14.7 (Nicolás 2007, Gerken 2008)**.** *When N is large enough, any N points in general position in the plane must contain an empty hexagon.*

It is well known that K_5 and $K_{3,3}$ are not planar. And a graph is planar if and only if it does not contain K_5 nor $K_{3,3}$ as a minor. The Euler's formula says, in a simple planar graph,

$$V - E + F = 2.$$

This easily implies

Fact 14.8. *If G is a planar graph with n vertices and m edges, then*

$$m \leq 3n - 6.$$

If $m > 3n - 6$, no matter how we embed the graph in the plane, there must be some *edge crossings*. But we may still want as few crossings as possible.

Definition 14.2. *Let G be a graph, the* crossing number $Cr(G)$ *is defined to be the least number of edge crossings for an embedding of G in the plane.*

Consider one embedding that achieves $Cr(G)$ crossings, we may find one crossing and remove one edge crossed, as long as there are more than $3n - 6$ edges. So

Fact 14.9. *For any graph G with n vertices and m edges,*

$$Cr(G) \geq m - 3n + 6.$$

The bound is fine if m is not too big. But it is very loose otherwise. Erdös and Guy conjectured in 1973 that when $m \geq 4n$, $Cr(G) \geq cm^3/n^2$ for some constant c. The first proof came in 1982 by Ajtai, Chvátal, Newborn, and Szemerédi, with $c = 1/100$; and independently by Leighton.

Theorem 14.10 (The crossing lemma)**.** *If G is a graph with n vertices and m edges, and $m \geq 4n$, then*

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

It was originally called a theorem, until in the 1990s in one paper Székely applyed it to many hard geometry problems, thus upgraded it to a lemma.

Proof. (Chazelle-Sharir-Welzl) Consider an embedding of the graph that achieves $Cr(G)$ crossings. Let $0 < p < 1$ TBD. We randomly pick an induced subgraph by randomly independently pick each vertex with probability p. Define the r.v.s X be the number of selected vertices, Y the number of edges, and Z the number of edge crossings. By Fact 14.9, in any outcome, we have $Z - Y + 3X \geq 6$, so

$$6 \leq \mathsf{E}(Z - Y + 3X) = p^4 Cr(G) - p^2 m + 3pn.$$

Simple calculation show that when $p = 4n/m$, we have the desired result. □

Given points in the plane, if we connect two points when their distance is 1, we form a so called *unit distance graph*. There are many hard problems related to the distance graphs. For example, the famous question about *the chromatic number of the plane*

Open Problem 14.1. *What is the minimum number of colours for colouring all the points in the plane so that any two points with distance 1 have different colours?*

All we know is that the answer is between 4 and 7.
With axiom of choice, one can prove that the number equals the maximum chromatic number of the unit distance graphs on finite point sets. However,

Shelah and Soifer showed in 2003 that this is no longer true if, instead of axiom of choice, we assume other consistent axioms.
Given n points in the plane, we form the unit distance graph among them. How many edges can we have? If we take roughly a $\sqrt{n} \times \sqrt{n}$ grid, then there are roughly $2n$ unit distances edges. The same order of number of edges is achieved by points that form many regular triangles. It is also clear that we can draw the high dimensional cube in the plane to get $n \log n$ unit distances.

Conjecture 14.2 (Erdös 1946)**.** *For any n points in the plane, the number of unit distances among them is $n^{1+o(1)}$.*

Erdös proved that the number of unit distances is $O(n^{1.5})$. It went through a series of improvements. $o(n^{1.5})$ Józsa-Szemerédi 1973; $O(n^{1.44...})$ Beck-Spencer 1984, and then today's best known bound

Theorem 14.11 (Spencer-Szemerédi-Trotter 1984)**.** *For any n points in the plane, the number of unit distances among them is $O(n^{4/3})$.*

All the proofs for these improvements use sophisticated methods. And the proof below is in Székely's short paper that demonstrated the glory of the crossing lemma.

Proof. (Székely 1997) Draw a unit circle centered at each point x if there are more than 2 points on the circle. And if there are multiple arcs between two points, we only keep one of them. View the arcs as edges, this gives us a drawing of a simple graph A on n points and m edges, where m is at least $(u - n)/2$, where u is the number of unit distances. Since any two circles intersect at most twice, so the number of crossings in this drawing is at most $2n^2$. Now use the crossing lemma and we are done. □

The next few problems discuss points in the plane and the lines determined by them (i.e. lines passing through at least two of these points).

Definition 14.3. *A set of points S is called* magic condifuration *if we can assign positive numbers to each point such that the sum on every line determined by S is 1.*

Let us try to find some magic configurations.
(a). All points on a line, we can put whatever positive weights we want, as long as their sum is 1.

(b). The pencil, $n-1$ points on a line with weight $1/(n-1)$, and one outsider with weight $(n-2)/(n-1)$.
(c). Points in general position, each with weight $1/2$.
And this one might be a little bit harder to find.
(d). Pick a triangle, $1/4$ on its vertices, $1/2$ on the middle point of each edge, and $1/4$ on its center.
The following was conjectured by Murty in 1971, and solved as late as 2007.

Theorem 14.12 (Ackerman-Buchin-Knauer-Pinchasi 2008)**.** *(a), (b), (c), and (d) above are all the magic configurations.*

The following problem was proposed by Sylvester in 1893, and solved 50 years later by Gallai.

Theorem 14.13 (Sylvester-Gallai theorem)**.** *n points in the plane. Either they are all on the same line, or there is a line that contains exactly two of the points.*

Proof. (Kelly) Consider a pair (l, v) where l is a line determined by the points, and v is a point outside l, and such that the distance from v to l is minimum. □

And 60 more years later, in 2003, the result was generalized in arbitary metric space. (You may think about what will be a "line" in metric space.)

Theorem 14.14 (de Bruijn-Erdös 1948)**.** *Given n points in the plane, either they are all on the same line, or they determine at least n lines.*

Proof. (1st proof) Use Sylvester-Gallai and induction. □

Proof. (2nd proof) Let $a_1, a_2, ..., a_n$ be the points, and $l1$, ..., l_m the lines. Let k_i be the number of points passing through a_i, and s_j be the number of points on l_j. (This can be viewed as the degrees in the bipartite graph that depicts the incidence relation of the points and lines. The idea is, if the k_i's are bigger than the s_j's, then there are more points on the l side.)
We observe that

$$\sum_i k_i = \sum_j s_j \tag{9}$$

and, if a point is not on a line with t points, then we can find at least t lines from that point

$$a_i \notin l_j \Rightarrow k_i \geq s_j \tag{10}$$

Now suppose a_n is the one that incident to fewest lines, i.e., $t := k_n$ is the smallest. W.L.O.G., it is incident to l_1, ..., l_t. We may number the points so that $a_1 \in l_1$, ..., $a_t \in l_t$. So we have, by (10)

$$k_1 \geq s_2, k_2 \geq s_3, ..., k_{t-1} \geq s_t, k_t \geq s_1 \tag{11}$$

And since k_n is the smallest and $a_n \notin l_q$ for $q > t$, so

$$k_q \geq k_n \geq s_q, \forall q > t. \tag{12}$$

(11), (12), together with (9) implies $m \geq n$. □

Proof. (3rd proof) Consider the $n \times m$ incidence matrix M where $M_{i,j} = 1$ if the point a_i is on the line l_j, otherwise 0. The matrix $A = MM^T$ has 1 off the diagonal (any two points see exactly one common line) and bigger than 1 on the diagonal (unless all the points are on the same line). It is easy to show that A has full rank, so the rank of M is at least n, and $m \geq n$. □

Now we look at the triples. Let us note down (with a family $\mathcal{F}$) all the (unordered) triples $\{a, b, c\}$ where a, b, and c are collinear. Then the line ab can be just defined as

$$\overline{ab} = \{a, b\} \cup \{c : \{a, b, c\} \in \mathcal{F}\}.$$

The property of the lines in the plane tells us that, for any four points, among its 4 triples, $\mathcal{F}$ can contain 0, 1, or 4, but not 2 nor 3.
More generally, we may first describe a family $\mathcal{F}$ of triples, and define the line as above. It could be read out of de Bruijn and Erdös' proof that if $\mathcal{F}$ has the property that, among any 4 points there are 0, 1, or 4 triples in $\mathcal{F}$, then either there is a line contains every point, or there are at least n different lines. Note that, now two "lines" may intersect on more than 2 points, and a line might be a proper subset of another.
In summer 2011 we found a generalization of this.

Theorem 14.15 (Beaudou - Bondy - Chen - Chiniforooshan - Chudnovsky - Chvátal - Fraiman - Zwols 2011)**.** *If $\mathcal{F}$ has the property that, among any 4 points there are 0, 1, 3, or 4 triples in $\mathcal{F}$, then either there is a line contains every point, or there are at least n different lines.*

Let us call "either there is a line contains everthing, or there are at least n lines" the dBE property. The original de Bruijn-Erdös says if we forbid 2 and 3 triples among every 4 points, then the configuration has dBE property. The generalized theorem says if we only forbid 2 triples among every 4, then the configuration has dBE property. What if we forbid 0 and 1 among 4? What if we forbid 0 and 3 among 4? We do not know yet.
On a slightly different direction, we are interested in the case where $\mathcal{F}$ is gotten from a finite metric space (X, ρ), where $\{a, b, c\} \in \mathcal{F}$ iff $\rho(a, b) + \rho(b, c) = \rho(a, c)$. Equivalently, $\mathcal{F}$ is generated by a connected weighted graph with positive lengths on edges such that $\{a, b, c\} \in \mathcal{F}$ iff $d(a, b) + d(b, c) = d(a, c)$. Here is the big conjecture

Conjecture 14.3 (Chen-Chvátal 2008)**.** *If $\mathcal{F}$ is generated by a metric space with n points, then either there is a line contains all the points, or there are at least n lines.*

For special cases we may consider the metric spaces generated by special graphs. It will be great if one could prove the conjecture for all the $\mathcal{F}$ generated by connected unit weights graphs.

15 Algebraic methods in combinatorics

How many subsets of $[n]$ can one find, such that the intersection of any two of them have the same size $t > 0$?

Example 15.1. *$t = n - 2$, and we pick all the elements in $\binom{[n]}{n-1}$. We have n sets.*

Example 15.2. *$A_i = \{i, n\}$, for $1 \leq i < n$. And $A_n = [n-1]$. $t = 1$. This can be viewed as a "pencil" configuration ($n-1$ collinear points and one outsider) in $\mathbf{R}^2$.*

Example 15.3. *The Fano plane. $n = 7$, there are 7 hyperedges, any two intersect at $t = 1$ point. Any projective plane of order n gives n sets with $t = 1$.*

In all these examples, the number of sets equals the number of points n. The following inequality tells that we cannot have more. It is called the *(non-uniform) Fisher's inequality.* [9]

Theorem 15.1. *Let A_1, A_2, ..., A_m be m distinct subsets of $[n]$ such that $|A_i \cap A_j| = t > 0$ for any $i < j$. Then $m \leq n$.*

Proof. If there is one set, say, A_1, such that $|A_1| = t$, then $A_1 \subseteq A_i$, and $A_i' = A_i - A_1$ are all disjoint. It is cleay in this case $m \leq n$. For the rest of the proof we assume $|A_i| > t$ for all i.
Consider the $m \times n$ *incidence matrix* M where $M_{ij} = 1$ if $j \in A_i$, 0 otherwise. Consider the $m \times m$ matrix $Q = MM^T$. Q_{ij} is the inner product of the i-th row of M and its j-th row, so $Q_{ij} = |A_i \cap A_j|$. Q has a t everywhere off-diagonal, and all the diagonal entries are bigger than t. We claim $r(Q)$, the rank of Q, is m, therefore the rank of M cannot be less than m, and $n \geq m$. There are various ways to prove that the matrix has full rank. One can compute the determinant directly, prove there is no non-zero solution to $Qv = 0$, or observe that Q is positive definite. □

[9] Just to mention that there are more examples where t is not 1 nor $n-2$. Let X be $\mathbf{F}_q^r$, the r dimensional space over $\mathbf{F}_q$. There are $n = (q^r - 1)/(q-1)$ one dimensional subspace of X, as well as n $(n-1)$-dimensional subspaces. Now let $A_i \subseteq [n]$ be the set of j's such that the j-th one dimensional is a subspace of the i-th $(n-1)$-dimensional subspace. Here each A_i is of size $(q^{r-1} - 1)/(q-1)$, and $t = (q^{r-2} - 1)/(q-1)$.

The uniform Fisher's inequality appeared in 1940. The proof is due to Majumder 1953, but can be essentially read out of Bose 1949.

Suppose $n = 100$, A_1, A_2, ..., A_m distinct subsets of $[n]$, such that each $|A_i|$ is even and $|A_i \cap A_j|$ is even for all $i < j$. How big can m be?
Group [100] into 50 pairs. For each pair, either pick none, or pick both. We have $m = 2^{50}$.
What if we change the rule a little bit? $|A_i \cap A_j|$ is even, but now each $|A_i|$ is odd. Let $A_i = \{i\}$, we get $m = n$. In fact there are many ways we can achieve $m = n$, but can we do better?

Theorem 15.2. *A_1, A_2, ..., A_m distinct subsets of $[n]$, each $|A_i|$ is odd and $|A_i \cap A_j|$ is even for all $i < j$, then $m \leq n$.*

Proof. Let M be the $m \times n$ incidence matrix. Now consider this as a matrix over $\mathbf{F}_2$. By the condition of this problem, $Q = MM^T$ is the identity matrix, therefore has rank m over the field $\mathbf{F}_2$, so M must have rank m and $n \geq m$.[10] □

Note: In retrospect, the statement of the problem indeed taste like Theorem 15.1 in a modulo form.

Theorem 15.3. *A_1, A_2, ..., A_m distinct subsets of $[n]$, each $|A_i|$ is even and $|A_i \cap A_j|$ is even for all $i < j$, then $m \leq 2^{\lfloor n/2 \rfloor}$.*

Proof. Suppose A_1, ..., A_m is a maximum collection. We identify a set $A_i \subseteq [n]$ with its indicator vector v_i in $\mathbf{F}_2^n$, a 0-1 vector of length n. The condition of this problem is equivalent to

$$(v_i, v_j) = 0, \forall i, j.$$

Let W be the subspace spanned by the v_i's, it is easy to see any $u, v \in W$ satisfies $(u, v) = 0$. Since m is maximal, so the v_i's fill the whole subspace and $m = 2^{\dim(W)}$.
On the other hand, the condition also tells that every v_i is in the kernel (a.k.a. null space) of W. So $\dim(W)$ is no more than the dimension of its kernel. The proof is finished by the fact that the dimension of the kernel is always $n - \dim(W)$. □

[10] Shang Jingbo [2011] noted that it is easy to see the determinant of Q in the proof above is odd, since there is exactly one odd number in the expansion.

If we want to break K_n into disjoint pieces, and each piece must be a complete bipartite graph. What is the least number of pieces possible? It is easy to see $n-1$ pieces is possible: We can use $K_{1,n-1}$ to reduce the problem to K_{n-1}. Or just use $K_{a,n-a}$ and recurse on K_a and K_{n-a}. Can we do better?

Theorem 15.4 (Graham - Pollak 1972)**.** *If the edge set of K_n is decomposed as the disjoint union of the edges sets of m complete bipartite graphs, then $m \geq n-1$.*

Proof. (Tverberg 1982) Let the m bipartite graphs be on the vertices sets (A_k, B_k), $1 \leq k \leq m$. Define

$$M := \sum_{k=1}^{m} \left(\sum_{i \in A_k} x_i \sum_{j \in B_k} x_j \right) = \sum_{k=1}^{m} \left(\sum_{i \in A_k, j \in B_k} x_i x_j \right) = \sum_{1 \leq i < j \leq n} x_i x_j.$$

So $\sum_{i=1}^{n} x_i^2 + 2M = (\sum_{i=1}^{n} x_i)^2$ (*). Now, if $m \leq n-2$, then the system of $m+1$ linear equations

$$\sum_{i \in A_k} x_i = 0 \quad k = 1, 2, ..., m$$
$$\sum_{1 \leq i \leq n} x_i = 0$$

has a solution that is not all 0. Contradicts (*). □

Proof. (In the matrix form) Let the m bipartite graphs be on the vertices sets (A_k, B_k), $1 \leq k \leq m$. Define the matrix M_k to be the 0-1 matrix where the (i,j) entry is 1 iff $i \in A_k$ and $j \in B_k$. So each M_k is just a rectangle with rank 1. Define $M := \sum_{i=1}^{m} M_k$. On one hand (exercise: for any matrices $r(A+B) \leq r(A) + r(B)$)

$$r(M) \leq \sum r(M_k) = m \quad (*).$$

On the other hand, since each edge of K_n is covered exactly once, so M is the adjacency matrix of a tournament. Therefore $M^T + M = J - I$. If $r(M) < n-1$, we get there is a non-zero solution x such that $Mx = 0$ and $\sum x_i = 0$. Thus

$$0 = x^T(M^T + M)x = x^T J x - x^T I x = -x^T x \neq 0,$$

a contradiction. So we must have $r(M) \geq n-1$, and by (*) $m \geq n-1$. □

The following *(upper) triangular criterion* is often applied with polynomials.

Lemma 15.5. *Let f_1, f_2, ..., f_m be functions $X \to \mathbf{F}$, and x_1, x_2, ..., $x_m \in X$. If*

$$f_i(v_j) \begin{cases} \neq 0 & \text{if } i = j \\ = 0 & \text{if } j < i \end{cases}$$

then the f_i's are linearly independent in the space $\mathbf{F}^X$.

It is clear we have the same conclusion if $j < i$ is replaced by $j > i$, in which case we call it lower triangular criterion.

Proof. Suppose there are λ_i's, not all 0, such that

$$f := \lambda_1 f_1 + \lambda_2 f_2 + ... + \lambda_m f_m \equiv 0.$$

Let t be the first such that $\lambda_t \neq 0$, we have

$$f(a_t) = \lambda_t f_t(a_t) + ... + \lambda_m f_m(a_t) = \lambda_t f_t(a_t) \neq 0,$$

a contradiction. □

A set of points in $\mathbf{R}^n$ is called a *unit distance set* if the distance between any two of them is a constant δ. It is well known we can have at most $n+1$ such points, they form a regular simplex.
What if two distances are allowed? One can construct a set of order n^2 such points (exercise). And we are going to prove that the bound is tight.

Theorem 15.6. *Let a_1, a_2, $a_m \in \mathbf{R}^n$, and $\delta_1, \delta_2 \in \mathbf{R}$. If for any $i \neq j$,*

$$\|a_i - a_j\| \in \{\delta_1, \delta_2\},$$

then $m \leq (n+1)(n+4)/2$.

Proof. Define, for each $1 \leq i \leq m$, a polynomial $\mathbf{R}^n \to \mathbf{R}$

$$f_i(x) = (\|x - a_i\|^2 - \delta_1^2)(\|x - a_i\|^2 - \delta_2^2).$$

Expanding the polynomials, it is easy to see each f_i is a linear combination of the following $(n+1)(n+4)/2$ polynomials

$$(\sum x_i^2)^2,\ (\sum x_i^2)x_j,\ x_i x_j,\ x_i,\ 1.$$

So we have $m \leq (n+1)(n+4)/2$ if the f_i's a linearly independent. This is indeed true by taking $v_i = a_i$ and apply the triangular criterion. □

Definition 15.1. *Let $L \subseteq [n] \cup \{0\}$. A hypergraph $\mathcal{F}$ on $[n]$ is* L-intersecting *if $|A \cap B| \in L$ for any $A \neq B$, $A, B \in \mathcal{F}$.*

When $|L| = 1$, we get the setting of Fisher's inequality. If, in general, $|L| = s$, how big can $\mathcal{F}$ be? One example is that when $L = \{0, 1, ..., s-1\}$, we may have

$$\mathcal{F} = \{A \in [n] : |A| \leq s\}. \ |\mathcal{F}| = \sum_{i=0}^{s} \binom{n}{i}.$$

The following is called *non-uniform Ray-Chaudhuri - Wilson theorem.*

Theorem 15.7. *(Frankl - Wilson 1981) If $|L| = s$ and $\mathcal{F} \in 2^{[n]}$ is L-intersecting, then $|\mathcal{F}| \leq \sum_{i=0}^{s} \binom{n}{i}$*

Proof. Let $L = \{l_1, l_2, ..., l_s\}$ and the sets (hyperedges) in $\mathcal{F}$ be A_1, ..., A_m, where

$$|A_1| \leq |A_2| \leq ... \leq |A_m|.$$

For each A_i, we associate an incidence vector v_i of length n

$$v_{i,j} = \begin{cases} 1 & \text{if } j \in A_i \\ 0 & \text{if } j \notin A_i \end{cases}$$

It is clear that the inner products $(v_i, v_j) = |A_i \cap A_j|$. Define the polynomials $\mathbf{R}^n \to \mathbf{R}$

$$f_i(x) = \prod_{k: l_k < |A_i|} ((v_i, x) - l_k)$$

Check that

$$f_i(v_j) \begin{cases} \neq 0 & \text{if } j = i \\ 0 & \text{if } j < i \end{cases} \qquad (*)$$

Moreover, we do the *multilinearization*, get $\widehat{f_i}$ from f_i by replacing each x_k^t with x_k, where $t > 1$. It is clear $\widehat{f_i}$ and f_i takes the same value on $\{0,1\}^n$, so (*) still holds with f_i's replaced by $\widehat{f_i}$'s. By the triangular criterion, the $\widehat{f_i}$'s are linearly independent; and they are in the subspace spanned by

$$\{\prod_{j \in T} x_j : |T| \subseteq [n], |T| \leq s\}.$$

□

For a graph (or a digraph) G, its adjacency matrix $A = A(G)$ is a natural algebraic subject to study. The (i, j) entry of A indicates whether there is an edge between i and j, i.e., in how many ways we can get from i to j in one step. By the definition of matrix multiplication, one can easily verify that A^k is the matrix where the (i, j) entry is the number of length k walks from i to j. For a matrix, i.e., a linear operator, its spectrum is certainly one thing we want to look at.
For graphs G of order n, its adjacency matrix A is an $n \times n$ real symmetric matrix. A can be expressed as $A = QDQ^{-1}$, where D is a diagonal matrix with n real eigenvectors λ_1, λ_2, ..., λ_n on the diagonal, and Q consists of n eigenvectors that form an orthonormal basis of $\mathbf{R}^n$. If there is no self-loop in G, the trace of A, which equals the sum of its eigenvalues, is 0. $A^k = QD^kQ^{-1}$, so it shares the same eigenvectors and has eigenvalues λ_i^k. Just take one step further, we have

Fact 15.8. *Let G be a un-directed graph (with possible parallel edges and self-loops), and $A(G)$ be its adjacency matrix. Suppose the eigenvalues of A are λ_1, ..., λ_n.*
Let P be a single variable polynomial. Then $P(A)$ has eigenvalues $P(\lambda_1)$, ..., $P(\lambda_n)$.

We list some other easy facts.

Fact 15.9. *Let A be the adjacency matrix of a simple graph G, then the trace of A is 0.*

Fact 15.10. *If G is a r-regular graph, then r is an eigenvalue of its adjacency matrix $A(G)$.*

Proof. It is easy to check, let v be the all 1 vector, $Av = rv$. □

Consider the graph with $2s + 1$ vertices, where one vertex has degree $2s$ and the others form a matching with s edges. In this graph every pair of vertices share exactly one common neighbour. This is called a *windmill graph.*

Theorem 15.11 (Erdös - Rényi - Sós 1966)**.** *If in G any two vertices has exactly one common neighbour, then G must be a windmill graph.*

Proof. Suppose there are n vertices. If there is one vertex with degree $n-1$, i.e., it is adjacent to every other vertex, then it is easy to show the graph is indeed a windmill.

Now suppose it is not the case, every degree is less than $n-1$. It is left as exercises to show that (1) If $u \not\sim v$, the $d_u = d_v$; and then (2) G is k-regular and $n = k^2 - k + 1$.
Let A be the adjacency matrix of G. For any $i \neq j$, there is exactly one common neighbour to them, i.e., there is exactly one walk of length 2 from i to j. Since each degree is k, there are exactly k walks from i, come back to i after 2 steps. So, $A^2 = J + (k-1)I$. We know that J has an eigenvalue n, and, since the rank of J is 1, its null space is of dimension $n-1$, so J has eigenvalues 0 of multiplicity $n-1$ and $n = k^2 - k + 1$ of multiplicity 1. So the eigenvalues for A^2 are $n-1$ $k-1$'s and one k^2. Thus A has one eigenvalue $\lambda_1 = k$, and all the others $|\lambda_i| = \sqrt{k-1}$. Note that A has trace 0, so $k = t\sqrt{k-1}$ for some integer t. It is easy to see this is only possible for $k = 2$, in which case $n = 3$ and the graph is windmill. □

Conjecture 15.1 (Kotzig 1979)**.** *For $l > 2$, there is no graph G such that, for any two vertices u and v, there is exactly one path of length l between u and v.*

Suppose G is a r-regular graph, and the girth of G is 5. Pick any vertex v, we draw the neighbours and second degree neighbours as a tree rooted at v. There are altogether $r^2 + 1$ vertices. Because there are no C_3 nor C_4, all these vertex are distinct. So, such graphs much have at least $n \geq r^2 + 1$ vertices. We are interested in the question: When $n = r^2 + 1$ is enough? In this case, the above picture, from any v, contains exactly all the vertices. So the distance between any two vertices is at most 2.
There are examples for $r = 2$ (the C_5), $r = 3$ (the Petersen graph). With a little work one may show there is no such graph with $r = 4$ and $n = 17$. It turns out there are no such graphs for $r = 5$ and $r = 6$. In 1960, Hoffman and Singleton gave such a graph with $r = 7$ ($n = 50$), and they proved that there is only one other possible case where $r = 57$ ($n = 3250$). [11]

Theorem 15.12 (Hoffman - Singleton 1960)**.** *If G is an r-regular graph with $n = r^2 + 1$ vertices, and girth 5, then $r \in \{2, 3, 7, 57\}$.*

Proof. Let A be the adjacency matrix of G. Follow the comment above, if $u \sim v$, then they don't have a common neighbour (G is C_3-free), so there is no walk of length 2 between them. If $u \not\sim v$, they are of distance 2, and

[11] After more than half a century, at the point of this writing, it is not known whether such a graph exist.

there cannnot be two walks of length 2 between them (G is C_4-free). So we may conclude that $A^2 + A = J + (r-1)I$. The eigenvalues for $J + (r-1)I$ is $r^2 + r$ with multiplicity 1, and $r - 1$ with multiplicaty $n - 1$. We know r is an eigenvalue of A. The other eigenvalues of A satisfy $\lambda^2 + \lambda = r - 1$. Let $s = \sqrt{4r-3}$. Suppose the eigenvalues $\frac{-1+s}{2}$ has multiplicity n_1, and $\frac{-1-s}{2}$ has multiplicity $r^2 - n_1$. Since A has trace 0, so

$$r + n_1 \frac{-1+s}{2} + (r^2 - n_1)\frac{-1-s}{2} = 0$$

and we get $r^2 - 2r = ts$ (*) for some integer t ($t = 2n_1 - n^2$). Either $t = 0$ and we get $r = 2$, or s is a positive integer, we get, from (*) and $r = (s^2+3)/4$,

$$s^4 - 2s^2 - 16ts - 15 = 0.$$

This says, s is a divisor of 15, so $s \in \{1, 3, 5, 15\}$. Note $r = (s^2+3)/4$. The case $s = 1$ gives $r = 1$ and $n = 2$, where the girth is not 5. For the others, $r \in \{3, 7, 57\}$. □

The following problem is from the 2007 International Mathematical Olympiad.

Example 15.1. *Let n be a positive integer. Consider*

$$S = \{(x, y, z) : x, y, z \in [n] \cup \{0\}, x + y + z > 0\}$$

as a set of $(n+1)^3 - 1$ points in $\mathbf{R}^3$. Determine the smallest number of planes, the union of which contains S but does not inlcude $(0, 0, 0)$.

It is easy to see $3n$ planes are doable. Actually there are many ways to do this: (1) Take the planes $x = 1$, $x = 2$, ..., $x = n$, $y = 1$, ..., $y = n$, $z = 1$, ..., $z = n$. (2) Take the planes $x + y + z = 1$, ..., $x + y + z = 3n$. (3) Some combinations of (1) and (2). It is not unreasonable to believe that the answer is $3n$.

The corresponding 2d problem, where we have a square instead of a cube, has a simple elementary proof: First we consider all the horizontal and vertical lines that are taken, then the rest un-covered points still form a 2d array of points. We can focus on the "boundary" of the array, and since no more horizontal or vertical lines, each line will intersect with the boundary at most twice.

Hint. Each plane gives a linear polynomial of the form $a_i x + b_i y + c_i z + d_i = 0$, consider the product of all these polynomials.

More hint. To keep the illustration managable, let us look at the algebraic method on the simplest case in 2d. Where we have 4 points $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$. We inspect some solutions to cover the 3 points with 2 lines.
1. $x=1$ and $y=1$. $(x-1)(y-1) = xy - x - y + 1$.
2. $x+y=1$ and $x+y=2$. $(x+y-1)(x+y-2) = 2xy+x^2-3x+y^2-3y+2$.
3. $x+y=1$ and $y=1$. $(x+y-1)(y-1) = xy - x + y^2 - 2y + 1$.
If you happen to imagine y^2 to be the same as y and x^2 to be x, then these 3 polynomials are essentially the same. This is not an accident. If $x=0$ or $x=1$, indeed $x^2 = x$.

You may know the fact that if a polynomial of degree at most n has $n+1$ roots, then it must be the zero polynomial. The following multi-dimensional generalization is (should be) also well known, and can be proved inductively from the 1d case. We omit the proof.

Lemma 15.13. *If $P(x,y,z)$ is a polynomial in x, y, z, and each term has degree at most (n,n,n). And suppose that there are $x_1 < x_2 < ... < x_{n+1}$, $y_1 < y_2 ... < y_{n+1}$, and $z_1 < ... < z_{n+1}$ such that $P(x_i, y_j, z_k) = 0$ for all i,j,k, then P is the zero polynomial.*

Consider $P - Q$, we have

Corollary 15.14. *If $P(x,y,z)$ and $Q(x,y,z)$ are two polynomials with degrees at most (n,n,n), and if $P(x,y,z) = Q(x,y,z)$ on all $0 \le x,y,z \le n$, then they are the same polynomial.*

Now we prove that in our problem, we need at least $3n$ planes.

Proof. Suppose we have a solution with t planes, the i-th plane is $a_i x + b_i y + c_i z + d_i = 0$. Define

$$P(x,y,z) := \prod_{1 \le i \le t} (a_i x + b_i y + c_i z + d_i).$$

Because each point is covered by at least one of the planes, and $(0,0,0)$ is not. So $P(x,y,z) = 0$ for all $(x,y,z) \in S$, and $P(0,0,0) \neq 0$.
Consider

$$R(x) = x^{n+1} - x(x-1)(x-2)...(x-n)$$

It is a polynomial of degree less than $n+1$, and $R(x) = x^n$ for all $0 \le x \le n$. Now we start from P and do a sequence of operations. Whenver there is a term $x^a y^b z^c$ where $a > n$, we replace x^{n+1} by $R(x)$.Formally, $x^a y^b z^c \to$

$x^{a-n+1}R(x)y^bz^c$. By doing this, the degree of x is reduced by at least 1. And importanly, let the resulting polynomial be P', we have $P(x,y,z) = P'(x,y,z)$ for all $0 \leq x,y,z \leq n$.
We keep doing this, and similarly for y and z. It is easy to see, after finite number of steps, we get a polynomial P^* such that
(a). $P^*(x,y,z) = P(x,y,z)$ for all $0 \leq x,y,z \leq 0n$.
(b). The degree of P^* is at most (n,n,n).
Now we plug in our trivial solution, let

$$Q(x,y,z) := (x-1)(x-2)...(x-n)(y-1)...(y-n)(z-1)...(z-n).$$

Q also vanishes on $0 \leq x,y,z \leq n$ except $(0,0,0)$, and Q also has degrees at most (n,n,n). So, after multiplying a constant, $P^* = cQ$ on $\{0,1,...,n\}^3$. By the lemma, $P^* \equiv cQ$ as polynomials.
Observe that the highest degree term in Q is $x^ny^nz^n$, so P^* has that term. On the other hand, if in the beginning we have $t < 3n$ planes, each terms $x^ay^bz^c$ has $a+b+c < 3n$. And clearly the reduction will not create any term that increases $a+b+c$, so $x^ny^nz^n$ will not appear in P^*, a contradiction. □

The whole proof indeed follows the outline of the well known

Theorem 15.15 (Combinatorial Nullstellensatz)**.** *Let $f(x_1,...,x_n)$ be a polynomial over a field $\mathbf{F}$. Suppose that the degree of f is t, and the coefficient of the monomial $x_1^{t_1}...x_n^{t_n}$ is nonzero, where $\sum_i t_i = t$. If S_1, ..., S_n are finite subsets of $\mathbf{F}$ with $|S_i| > t_i$ for each i, then there are $s_i \in S_i$ such that*

$$f(s_1,...,s_n) \neq 0.$$

We omit the proof, but comment that the proof is essentially presented in the previous example, the key step is replacing a high degree term by many lower degree ones. Alternatively, if one uses Combinatorial Nullstellensatz, we do not need to go through the reduction in the previous problem.
Among the numerous applications, Combinatorial Nullstellensatz gives a simple proof to the following classical theorem in number theory.

Theorem 15.16 (Cauchy-Davenport)**.** *Let p be a prime, A and B be two non-empty subsets of $\mathbf{Z}_p$, then*

$$|A+B| \geq \min\{p, |A|+|B|-1\}.$$

Remark. It is easy to see the bound is tight. For example, take A to be the lowest s elements in $\mathbf{Z}_p$, and B be the lowest t elements. The theorem is quite easy when $|A| + |B| > p$. Also note that the statement is not true when p is not a prime.

Proof. We may assume $|A| = s$, $|B| = t$, and $s+t \leq p$, otherwise the theorem is trivial. Suppose $|A + B| \leq s + t - 2$, pick a set C such that $A + B \subseteq C$ and $|C| = s + t - 2$. Let

$$f(x, y) = \prod_{c \in C}(x + y - c).$$

On one hand, f vanishes on (A, B). On the other hand, it is of degree $s+t-2$, and the coefficient of $x^{s-1}y^{t-1}$ is $\binom{s+t-2}{s-1}$, which is not 0 in $\mathbf{Z}_p$ when $s + t \leq p$. This contradicts with the Combinatorial Nullstellensatz with $n = 2$, $S_1 = A$, and $S_2 = B$. □

We repeat the theorem $k - 1$ times, get

Corollary 15.17. *Let p be a prime and A_1, ..., A_k be non-empty subsets of $\mathbf{Z}_p$, then*

$$|A_1 + ... + A_k| \geq \min\{p, \sum_i |A_i| - (t - 1)\}.$$

Theorem 15.18 (Erdös-Ginzburg-Ziv 1961)**.** *For any positive integer n, and any $2n - 1$ integers a_1, a_2, ..., a_{2n-1}, there are always n of them whose sum is a multiple of n.*

We left it as an exercise to show that, if the statement in the theorem holds for numbers n_1 and n_2, then it holds for $n = n_1n_2$. So it suffice to show that the statement holds for any prime number. We need to prove the following

Theorem 15.19. *For any prime number p, and any $2p - 1$ numbers a_1, a_2, ..., a_{2p-1}, there are always p of them whose sum is a multiple of p.*

Proof. (Original proof) If there is a number repeated at least p times among the a_i's, then their sum is already a multiple of p. Otherwise, we assume

$$a_1 \leq a_2 \leq ... \leq a_{2p-1}.$$

Let

$$A_1 = \{a_1, a_{p+1}\},\ A_2 = \{a_2, a_{p+2}\}, ..., A_{p-1} = \{a_{p-1}, a_{2p}\}, A_p = \{a_p\}.$$

Now we have $a_t \neq a_{p+t}$, so $|A_i| = 2$ for $i < p$. Now apply Corollary 15.17. □

Proof. Consider the sum

$$S := \sum_{I \in \binom{[2p-1]}{p}} \left(\sum_{i \in I} a_i \right)^{p-1}.$$

If none of the p subsets has 0 sum in $\mathbf{Z}_p$, by Fermat's little theorem and Lucas theorem,

$$S \equiv \binom{2p-1}{p} \not\equiv 0.$$

On the other hand, any term in the expansion has the form

$$a_{s_1}^{t_1} ... a_{s_k}^{t_k} : k < p, \sum_i t_i = p - 1.$$

Its coefficient is a multiple of

$$\binom{2p-1-k}{p-k} \equiv 0,$$

by Lucas again. A contradiction. □

Let (A, B) be a complete bipartite graph, with $A = \{a_1, a_2, ..., a_n\}$ and $B = \{b_1, b_2, ..., b_n\}$. Suppose there is a weight $w_{i,j}$ on the edge $a_i b_j$. Naturally we associate it with a matrix M where the (i, j) entry is $w_{i,j}$. The determinant of M is

$$\det M = \sum_\sigma \text{sign}\sigma \prod_i w_{i,\sigma(i)}.$$

Each term arises from a permutation σ, which can be viewed as a perfect matching between A and B, or a max flow if all the vertex capacities are 1, or let us call it a maximal family of disjoint paths.
Let G be a directed acyclic graph with weights w_e on each edge e. Extend the weights to paths such that for any directed path P,

$$w(P) := \prod_{e \in P} w(e).$$

Natually $w(P) = 1$ if P has length 0. For a path system $\mathcal{P} = \{P_1, P_2, ..., P_t\}$, we also define

$$w(\mathcal{P}) := \prod_{P \in \mathcal{P}} w(P).$$

For any two vertices u and v, define

$$w(u,v) := \sum_{P:u\to v} w(P).$$

Note that $w(u,v) = 0$ if there is no path from u to v, and $w(u,u) = 1$.
For two vertex sequences $A = (a_1, ..., a_n)$ and $B = (b_1, ..., b_n)$, let us call $\mathcal{P} = \{P_1, P_2, ..., P_n\}$ an A-B *(matching) family* if there is a permutation σ such that P_i is from a_i to $b_{\sigma(i)}$. Define the sign of $\mathcal{P}$ to be $\text{sign}\mathcal{P} := \text{sign}\sigma$. An A-B (matching) family $\mathcal{P}$ is called *vertex-disjoint* if all the paths in $\mathcal{P}$ are vertex disjoint. Denote $\mathcal{F}(A,B)$ the set of all A-B families, and $\mathcal{F}^*(A,B)$ the set of all vertex disjoint A-B families.

Lemma 15.20 (Lindström 1972, Gessel - Viennot 1985)**.** *Let $G = (V,E)$ be a finite directed acyclic graph with weight function $w : E \to \mathbf{R}$, $A = (a_1, a_2, ..., a_n)$, $B = (b_1, b_2, ..., b_n)$ where $a_i, b_i \in V$ (no other restrictions, the same vertex can appear anywhere multiple times). Let M be the $n \times n$ matrix where the (i,j) entry is $w(a_i, b_j)$. Then*

$$\det M = \sum_{\mathcal{P}\in\mathcal{F}^*(A,B)} sign(\mathcal{P})w(\mathcal{P}).$$

Note: If G is directed from a bipartite graph (A,B), with all the edges from A to B, then the lemma is indeed just what we discussed about the determinant of any square matrix.
Note: If $a_i = a_j$, then $\mathcal{F}^*(A,B) = \emptyset$, and $\det M$ is indeed 0 since there are two identical rows.
Note: The lemma is almost "obvious" in the sense that it is easy to see that

$$\det M = \sum_{\mathcal{P}\in\mathcal{F}(A,B)} \text{sign}(\mathcal{P})w(\mathcal{P}).$$

So, let $\mathcal{G}(A,B) := \mathcal{F}(A,B) - \mathcal{F}^*(A,B)$ be the set of all the A-B families where some paths intersect, all we need to show is that

$$\sum_{\mathcal{P}\in\mathcal{G}(A,B)} \text{sign}(\mathcal{P})w(\mathcal{P}) = 0.$$

Proof. We define a mapping $f : \mathcal{G} \to \mathcal{G}$. For any $\mathcal{P} \in \mathcal{G}$, let i^* be the smallest i such that P_i intersects some other path, j^* be the smallest j such that P_{i^*}

intersects P_j, and v be the first intersection of P_{i^*} and P_{j^*}. f keeps all the paths, except switching P_{i^*} and P_{j^*} at the point v.
It is easy to check $f(f(\mathcal{P})) = \mathcal{P}$, $w(\mathcal{P}) = w(f(\mathcal{P}))$, and $\text{sign}\mathcal{P} = -\text{sign}f(\mathcal{P})$. □

Exercise 15.1. *Let $a_1 < ... < a_n$ and $b_1 < ... < b_n$ be two sequence of positive integers. Let M be the matrix where the (i,j) entry is $\binom{a_i}{b_j}$. Then $\det M \geq 0$, and it is 0 iff $a_i < b_i$ for some i.*

The lemma gives a simple proof to the classical Cauchy-Binet formula in linear algebra.

Theorem 15.21 (Cauchy-Binet formula)**.** *If A is an $n \times m$ matrix and B an $m \times n$ matrix, then*

$$\det(AB) = \sum_{X \in \binom{[m]}{n}} \det(A_X) \det(B_X),$$

where A_X is the sub-matrix of A taking the columns in X, and B is the sub-matrix of B taking the rows in X.

Proof. Draw 3 rows of vertices. The first row $\mathcal{A} = (a_1, a_2, ..., a_n)$, the second row $\mathcal{C} = (c_1, c_2, ..., c_m)$, and the third $\mathcal{B} = (b_1, b_2, ..., b_n)$. Make it a complete bipartite graph between $\mathcal{A}$ and $\mathcal{C}$, as well as between $\mathcal{C}$ and $\mathcal{B}$. To make it a d.a.g., all the edges are from top down. $w(a_i c_j) = A_{i,j}$, and $w(c_i b_j) = B_{i,j}$. So far so natural.
Let $M = AB$, observe that $M_{i,j}$ is indeed $w(a_i, b_j)$. By the lemma, $\det M$ equals the sum of $\text{sign}\mathcal{P}w(\mathcal{P})$, where $\mathcal{P}$ ranges over all the vertex disjoint path family from $\mathcal{A}$ to $\mathcal{B}$. Note that any $\mathcal{A}$-$\mathcal{B}$ vertex disjoint path family must go though a set X of n vertices in the middle level. We group the sums by X, using the fact that $\text{sign}\sigma_1 \text{sign}\sigma_2 = \text{sign}(\sigma_1\sigma_2)$ and prove the partial sum is indeed $\det(A_X)\det(B_X)$.
For the last step, we could also "cheat" by saying that the $n = m$ case for the formula is the well known fact that $\det(AB) = \det(A)\det(B)$ for two square matrices. □

Consider a graph G with n vertices and m edges (with possible parallel edges, but no self loops), now we make $\overrightarrow{G}$, an arbitary orientation (i.e., give one direction to each edge). The incidence matrix Q of $\overrightarrow{G}$ is the $n \times m$ matrix

where, if the i-th edge is $\overrightarrow{u_i v_i}$, the i-th column has a 1 on the row u_i, a -1 on the row v_i, 0 everywhere else.
Let us observe the matrix $M := QQ^T$. On the diagonal it has $M_{i,i}$ the degree of the i-th vertex, and $-M_{i,j}$ is the number of edges between i and j. Notice that M does not depend on the orientations of the edges.

Definition 15.2. *Let G be a graph, the matrix M defined as above is called the* Laplacian matrix *for G.*

What can we say about the Laplacian matrix? It does not have full rank, simply because all the rows of Q add up to 0. Things gets interesting when we consider a $(n-1) \times (n-1)$ principle submatrix of M.

Fact 15.22. *Let Q_x be the matrix Q with its row x deleted. If e_1, e_2, ..., e_k form a cycle, then the corresponding k columns are linearly dependent.*

Proof. Walk along the cycle. Multiply the columns by -1 if needed, we can get a 1 on row t whenever we enters the vertex t, and -1 whenever we leave it. So the columns (with coefficients ± 1) add up to 0. □

And another nice observation.

Fact 15.23. *Let Q_x be the matrix Q with its row x deleted. If $T = \{e_1, e_2, ..., e_{n-1}\}$ form a tree, then the corresponding $n-1$ columns are linearly independent. In fact, let $Q_{x,T}$ be the $(n-1) \times (n-1)$ submatrix of Q_x, where we take all the rows except x, and all the columns corresponding to T, then*

$$\det Q_{x,T} = \pm 1.$$

Proof. Induction. Use the fact that there are at least two leaves in a tree. Expand the determinant by the row of a leaf. □

Now we reach the matrix-tree theorem.

Theorem 15.24 (Kirchhoff 1847, implicitly)**.** *Let $G = (V, E)$ be a graph on $[n]$ with possible parallel edges but without self-loops, let M be its Laplacian matrix and M_x be the submatrix of M by deleting the x-th row and x-th column. Then the number of spanning trees of G equals $\det M_x$ for any x.*

Proof. Notice that $M_x = Q_x Q_x^T$. By Cauchy-Binet we have

$$\det M_x = \det(Q_x Q_x^T) = \sum_{E' \in \binom{E}{n-1}} \det(Q_{x,E'}) \det(Q_{x,E'}^T) = \sum_{E'} \det(Q_{x,E'})^2.$$

By our observations above, any terms in the last sum is 1 if E' is a spanning tree, otherwise it is 0. □

In this section we calculated several times the eigenvalues of matrices of the form $J + kI$. The classical Cayley's formula immediately follows.

Theorem 15.25 (Cayley 1889)**.** *There are n^{n-2} different spanning trees in K_n.*

16 Turán graphs, extremal graph theory

What is the maximum number of edges a graph on n vertices can have, such that the graph is still triangle-free? More general question: What is the maximum number of edges a graph on n vertices can have, such that it is K_p free, for some fixed p? Suppose the maximum number is m, can we describe all the "extremal graphs" on n vertices and exactly m that are still K_{p+1}-free ($p \geq 1$)? Questions of this kind belong to the subject of *extremal graph theory*. "Given a property $\mathcal{P}$, what is the maximum (minimum) number of edges (sequence of degrees) such that a graph on n vertices having property $\mathcal{P}$? And what are the graphs that achieve the maximum (minimum)?"
For many questions one can ask the "extremal" version. Thus the extremal graph theory (or extremal combinatorics) may touch almost all the subjects in combinatorics. For example, all the Ramsey type questions (though they already grow to a big subject themselves) can be viewed as in this category, as well as the bounds we studied for property B. As we have seen, the exact bounds for those questions are hard to get. On the other hand, the two questions above admits easy solutions, they are completely solved, and are considered the usual starting point of the subject of extremal graph theory. This chapter serves as an introduction to the famous results, as well as a showcase of many different techniques in graph theory and combinatorics.
The answer to the first question was due to Mantel.

Theorem 16.1 (Mantel 1907)**.** *For any n, if G is a triangle-free graph on n vertices, then $|E(G)| \leq \lfloor n^2/4 \rfloor$.*

It is clear $K_{\lfloor n/2 \rfloor \lceil n/2 \rceil}$ is triangle-free and indeed has $\lfloor n^2/4 \rfloor$ edges. In fact it is the only extremal graph that are triangle-free and with $\lfloor n^2/4 \rfloor$ egdes on n vertices. Interested readers can prove this as an exercise. We mention that this is a special case of the more general Turán's theorem that we will prove later.

Proof. (of Theorem 16.1, orioginal proof) If G has no triangles, then for any edge uv, $d(u) + d(v) \leq n$, sum over all the edges,

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

For each u, $d(u)$ appears on the l.h.s. $d(u)$ times, so $\sum_{u \in V} d(u)^2 = nm$ (*). Now use Cauchy (which say, for two vectors $\overrightarrow{x}$ and $\overrightarrow{y}$, their inner product

is at most the product of their lengths),

$$4m^2 = \Big(\sum d(u)\Big)^2 \leq n \sum d(u)^2.$$

Plug this into (*) we get $m \leq n^2/4$. □

Proof. (of Theorem 16.1, 2nd proof) Let A be a largest independent set in G, and $B = V - A$. Since A is an independent set, each edge has at least one end in B. So $m \leq \sum_{v \in B} d(v)$ (*). On the other hand, since there is no triangle, $N(v)$ is an independent set for any v, so $d(v) \leq |A|$. Plug into (*), we have $m \leq |B||A| \leq n^2/4$. □

And here is the place I can taste my favorite method.

Proof. (of Theorem 16.1, 3rd proof, the shifting method) We assign a weight $w(v)$ to each vertex v, and define the variable $S = \sum_{uv \in E} w(u)w(v)$. Initially, $w(v) = 1/n$ for any v, so $S = m/n^2$. In each step, if there are u and v such that $uv \notin E$, $w(u) = w_1 > 0$ and $w(v) = w_2 > 0$ (*), let $S_1 = \sum_{x \sim u} w(x)$ and $S_2 = \sum_{x \sim v} w(x)$, w.o.l.g., $S_1 \geq S_2$, then we change $w(u)$ to $w_1 + w_2$ and $w(v)$ to 0. Thus, in each step, we have (1) $\sum_v w(v)$ is still 1; (2) number of non-zero w's is decreased by 1; and (3) S is not decreasing, we always have $m/n^2 \leq S$.
By (2), this process must end. When it ends, there are no u and v satisfying (*), i.e., all the v's where $w(v) > 0$ form a clique. But the graph is triangle-free, so all the weights are on a single edge uv, and by (1) $S = w(u)w(v) = w(u)(1 - w(u)) \leq 1/4$. □

We turn to the question of which graphs has no K_{p+1}.

Definition 16.1. *A graph $G = (V, E)$ is* p-partite *if V can be partitioned into $V_1, ..., V_p$ such that there are no edges inside each part. A p-partite graph is a* complete p-partite graph *if uv is an edge whenever u and v are from different parts.*

It is easy to see that there is no K_{p+1} in any p-partite graph. To get as many edges as possible, we make the graph complete, and whenever there are two parts such that $|V_i| - |V_j| > 2$, we move one point from V_i to V_j, the number of edges increases. We also note that there is only one complete p-partite graph on n vertices, up to isomorphism, such that $|V_i| - |V_j| \leq 1$ for any pairs (i, j).

Definition 16.2. *Suppose $n = tp + r$, where $0 \leq r < p$. An* equi-partition of n into p parts *is $(x_1, x_2, ..., x_p)$ such that $x_1 = ... = x_r = \lceil n/p \rceil$ and $x_{r+1} = ... = x_p = \lfloor n/p \rfloor$.*

Definition 16.3 (Turán graphs)**.** *For each n and p, the Turán graph $T(n, p)$ is the complete p-partite graph on n vertices where the sizes of the parts form an equi-partition of n. We denote the number of edges in $T(n, p)$ by $t(n, p)$.*

We note that $t(n, p)$ can be easily calculated.

Fact 16.2. *If $n \equiv r \mod p$, then*

$$t(n, p) = \frac{p-1}{2p}n^2 - \frac{r(p-r)}{2p}.$$

Theorem 16.3 (Turán 1941)**.** *Let G be a graph on n vertices and m edges. If $\omega(G) \leq p$, then $m \leq t(n, p)$. Moreover, if $m = t(n, p)$ and $\omega(G) \leq p$, then $G \cong T(n, p)$.*

Look at the other side of the world, we restate the theorem as follows

Theorem 16.4. *For any n and p, let $C_{n,p}$ be the graph with p cliques whose sizes form an equi-partition of n. Let G be a graph on n vertices and m edges. If $\alpha(G) \leq p$, then $m \geq |E(C_{n,p})|$. Moreover, if $m = |E(C_{n,p})|$ and $\alpha(G) \leq p$, then $G \cong C_{n,p}$.*

Note that the third proof for Mantel's theorem, which uses the weights and shifting the weights between vertices, almost works for Turán's theorem. The proof was due to Motzkin and Straus.

Proof. (of Theorem 16.3 by Turán) Prove by induction. The base case is for any $n \leq p$, the statement is clearly correct.
Now let $n > p$. We may assume G is a maximal graph without K_{p+1}, i.e., adding any edge creates a K_{p+1} in G. So there is a K_p in G, denote the p vertices of this clique by S. For any $v \in V - S$, there are at most $p-1$ edges from v to S, otherwise we find a K_{p+1}. So there are at most $(n-p)(p-1)$ edges between S and $V - S$. By the inductive hypothesis (when $n' = n - p$), there are $t(n-p, p)$ edges inside $V - S$. So,

$$m \leq \binom{p}{2} + (n-p)(p-1) + t(n-p, p).(*)$$

In parallel, look at any K_p in $T(n,p)$, we find the number of edges is indeed

$$t(n,p) = \binom{p}{2} + (n-p)(p-1) + t(n-p,p).$$

When the equality in (*) holds, by induction, $V-S$ induces $T(n-p,p)$, and there are exactly $(n-p)(p-1)$ non-edges between S and $V-S$. Each point in $V-S$ meets exactly one such non-edge, and each vertex v in S must have at least one part, out of the p-parts in $T(n-p,p)$, such that there is no edge from v to that part (Otherwise we find a K_{p+1}). It is easy to see that, when these conditions are satisfied, the graph is exactly $T(n,p)$. (The detailed are left for the reader to check, and be a little careful when $n-p<p$.) □

For the other proofs, we state a trivial observation.

Fact 16.5. *A graph is a disjoint union of cliques iff for any three points u, v, w, $uv \in E$ and $vw \in E$ implies $uw \in E$. A graph G is multi-partite iff for any three points u, v, w, $uv \notin E$ and $vw \notin E$ implies $uw \notin E$.*

Proof. (of Theorem 16.3 by Alon - Spencer) This time we prove the version of the theorem in its complement graph. Let $m^* := |E(C_{n,p})|$.
Consider a random ordering σ of V. For any $v \in V$, define the event "v is a seed" if $v \prec_\sigma u$ for any $uv \in E(G)$. And define X_v to be the indicator random variable for this event. Moreover, define $X_G := \sum X_v$ to be the number of seeds in an outcome. It is clear that all the seeds form an independent set, so

$$\alpha(G) \geq \max(X_G) \geq \mathsf{E}(X_G) = \sum_{v \in V(G)} \frac{1}{1+d_v},$$

with equality holds only if X_G is a constant over the σ's. It is also clear $X_{C_{n,p}}$ is indeed constantly p over all the permutations. By the convexity of the function $1/(1+x)$, the r.h.s. is at least $\sum \frac{1}{1+d'_v}$, where the d'_v form an equi-partition of $2m$. Note that the degrees of $C_{n,p}$ form an equi-partition of $2m^*$. So, when $m \leq m^*$, $\mathsf{E}(X_G) \geq \mathsf{E}(X_{C_{n,p}}) = p$. The equality holds only if $m = m^*$.
We are left to prove that $C_{n,p}$ is the only extremal graph. It is enough to show that the extremal graph is a disjoint union of cliques. Suppose there are $uv \in E$ and $vw \in E$ but $uw \notin E$. Let A be the event that (u,w,v) is the prefix of the random permutation σ, and B be the event that (u,v,w) is the prefix. We have $\mathsf{E}(X_G|A) = \mathsf{E}(X_G|B) + 1$, so X_G is not a constant. Follow our proof, either $m < m^*$, or $\alpha(G) > p$. □

Definition 16.4. *Two vertices v and v' are called* (non-adjacent) twins *in a graph G if $N(v) = N(v')$.*

Proof. (of Theorem 16.3) Start from G. As long as G is not multi-partite, there are $uv \notin G$, $vw \notin G$, but $uw \in G$.
Case 1. $d_u < d_v$ or $d_u < d_w$. W.l.o.g., $d_u < d_v$. Remove u, and make a twin for v, there are more edges, and still no K_{p+1}.
Case 2. $d_u \geq d_v$ and $d_u \geq d_w$. Remove both v and w, and make two twins for u, there are one more edge, and still no K_{p+1}.
So we end up with a s-partite graph with $s \leq p$, if it is still not $T(n,p)$, we change it to $T(n,p)$, there are more edges. □

We turn to the question about paths and cycles. It is more interesting to talk about degrees instead of total number of edges.

Theorem 16.6 (Dirac - Pósa)**.** *Let G be a connected graph on n vertices with $\delta(G) \geq k$. (i) If $2k \geq n$, then there is a Hamiltonian cycle in the graph. (ii) If $2k < n$, then G contains a path of length $2k+1$.*

Proof. (Sketch) Consider the longest path $P_l = (x_1, x_2, ..., x_l)$ in G. Since it is a longest path, all the neighbours of x_1 must be in the x_i's, so are all the neighbours of x_l. If $l \leq 2k$, then $\delta(G) \geq k$ will lead us to an index i such that $x_1 \sim x_i$ and $x_l \sim x_{i-1}$, and then we can find a cycle of length l. Now, if $l \leq 2k$ and $l < n$, since G is connected, there is one point adjacent to a point on the cycle, and we can find a path of length $l+1$, a contradiction. So, either $l > 2k$, or $l = n$. (i) and (ii) follows. □

Remark. (i) is Dirac's theorem, and later Pósa strengthened it with a theorem slightly better than the above. In the proof, it is enough to have $d_u + d_v \geq 2k$ for any u and v, but that is not much different than $d_u \geq k$ for all u.
In many applications the form above is good enough. We note that there are improvements to it. At the end of the line is

Theorem 16.7 (Chvátal)**.** *Let G be a graph with degree sequence $d_1 \leq d_2 \leq ... \leq d_n$, $n \geq 3$. If $d_{n-k} \geq n-k$ whenever $d_k \leq k < n/2$, then G has a Hamiltonian cycle.*

Here is a cute puzzle from the textbook.

Example 16.1. *A 3×3 cube of cheese is divided into 27 unit cubes. A mouse eats one unit cube each day and move to an adjacent unit (sharing a face) the next day. Can the mouse eat the center unit cube on the last day?*

Theorem 16.8. *G a graph on n vertices and more than $n\sqrt{n-1}/2$ edges, then the girth of G is at most 4. i.e., G contains a triangle or C_4.*

Proof. Suppose G has no C_3 nor C_4, for any vertex u, consider the vertices of distance 1 and 2 from u,

$$\sum_{v\sim u} d_v \leq n-1.$$

Sum over all u's, we get

$$\sum_{v} d_v^2 \leq n(n-1).$$

And apply Cauchy. □

Follow the proof, if there are indeed $n\sqrt{n-1}/2$ edges and C_3, C_4-free, then, (i) for any u, any other vertex is of distance 1 or 2 away from u; (ii) the equality in Cauchy holds, implies that all the degrees are the same, the graph is $\sqrt{n-1}$-regular. It is shown that this is possible for $n = 5$ (C_5), $n = 10$ (Petersen graph), $n = 50$ (Hoffman - Singleton graph), and the only other possible value is $n = 3250$.
Some more exercises:

Exercise 16.1. *If G is a graph on n vertices and does not contain C_4, then G has at most $n(1+\sqrt{4n-3})/4$ edges.*

Exercise 16.2. *If G is a graph on n vertices and $\delta(G) \geq (n+1)/2$, then for any two vertices u and v, there is a Hamiltonian path from u to v.*

Exercise 16.3. *If a simple graph on n vertices has m edges, then it has at least $m(4m-n^2)/3n$ triangles.*

Turán's theorem says if there are roughly more than $(1-\frac{1}{p})\frac{n^2}{2}$ edges, then we can find a K_{p+1}. The following theorem says if we add slightly more edges, then we can always find a copy of $T(q, p+1)$ for any fixed q.

Theorem 16.9 (Erdös - Stone 1947)**.** *For any $p \geq 2$, $q \geq 2$ and $0 < \epsilon$, there is an integer $N = N(p, q, \epsilon)$ such that any graph with $n \geq N$ vertices and $(1-\frac{1}{p}+\epsilon)\frac{n^2}{2}$ edges contains a copy of $T(q, p+1)$.*

Definition 16.5. *Let H be a graph. For any integer n, $ex(n, H)$ is defined to be the maximum number of edges among graphs on n vertices that does not contain a copy of H.*

Turán's theorem states that $ex(n, K_{p+1}) = t(n, p)$. Erdös-Stone (as in the form we stated it, together with the Turán graphs) implies that

$$ex(n, T(q, p+1)) \in \left(1 - \frac{1}{p} + o(1)\right) \frac{n^2}{2}.$$

We leave the proof of Erdös-Stone for next section. Note that when $t = p+1$ this is almost Turán's theorem. Yet it is not only a simple generalization of Turán's theorem. It is often called *the fundamental theorem of extremal graph theory*. One of its application is the surprising role of the chromatic number in extremal graph theory (vs. the role of number of edges in random graphs).

Theorem 16.10 (Erdös - Stone - Simonovits)**.** *If H is a graph with chromatic number $\chi(H) = p + 1$, then*

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

Proof. For the lower bound, consider the graph $T(n, p)$, it cannot contain H, otherwise $\chi(H) \leq p$.
For the upper bound, apply Erdös-Stone with $q = |V(H)|(p + 1)$. □

We end this section with the most famous conjecture about extremal problems.

Conjecture 16.1 (Erdös - Sós 1963)**.** *If G is a graph on n vertices and more than $n(k-1)/2$ edges, then G contains every tree with k edges.*

17 Szemerédi regularity lemma

Ramsey theorem (on graphs) asserts that, if the edges of the complete graph K_n is coloured by s colours, when n is big enough ($n > N_s(q;2)$), then we can always find a monochromatic K_q. The theorem does not tell in which colour class the K_q resides. Can it be in the largest colour class? e.g. If K_n is coloured by yellow and blue, where the number of yellow edges is more than $\binom{n}{2}/2$, can we always find a yellow K_q? The answer is negative according to the Turán graphs.
By contrast, consider van der Waerden theorem, if we colour $[n]$ with s colours, when n is big enough, we can always find a monochromatic A.P. with length t. In 1936, four years before Szemerédi was born, Erdös and Turán conjectured a stronger version: As long as the colour class has positive upper density, we can find an A.P. of this colour. Roth in 1956 proved the first non-trivial case $t = 3$ using analytical method. Szemerédi proved the case $t = 4$ in 1969 with combinatorial method. Finally,

Theorem 17.1 (Szemerédi 1975)**.** *For any $d > 0$, and any positive integer t, there exists $N = N(d,t)$ such that any subset $S \subseteq [N]$ where $|S| \geq dN$ contains an arithmetic progression of length t.*

It was in that paper Szemerédi introduced his lemma, which became one of the major tools in modern combinatorics.

Definition 17.1. *Let $G = (V,E)$ be a graph. X and Y two disjoint nonempty sets of V. Let $e(X,Y)$ be the number of edges between X and Y. The* density *of the pair (X,Y) is*

$$d(X,Y) = \frac{e(X,Y)}{|X||Y|}.$$

Pick a desired density $0 \leq d \leq 1$, if we randomly pick the edges, for each $(x,y) \in X \times Y$, xy becomes an edge with probability d. Then the expected density $\mathsf{E}(d(X,Y))$ is indeed d. Moreover, this is true for any nonempty subsets: $A \subseteq X$ and $B \subseteq Y$, then $\mathsf{E}(d(A,B)) = d$. In particular, for any vertex $x \in X$, the expected number of its neighbours in Y is $d|Y|$. It is also true that, with high probability, the values are close to their expected value.

Definition 17.2 (regular pair)**.** *Let $\epsilon > 0$ be a (small) constant. A pair (X,Y) is called* ϵ-regular *if, for any*

$$A \subseteq X, B \subseteq Y, |A| > \epsilon|X|, \text{ and } |B| > \epsilon|Y|,$$

we have

$$|d(A,B) - d(X,Y)| < \epsilon.$$

Otherwise it is called irregular.

The regular pair with density d resembles the random bipartite graph with parameter d in the following way. When B is a decent subset of Y, most of vertices in X has close to the expected $d|B|$ neighbours in B.

Fact 17.2. *Let (X,Y) be an ϵ-regular pair with $d(X,Y) = d$; let B be any subset of Y where $|B| > \epsilon|Y|$. Then*

$$|\{x \in X : |B \cap N(x)| \leq (d-\epsilon)|B|\}| \leq \epsilon|A|.$$

Proof. By contradiction. Let $A = \{x \in X : |B \cap N(x)| \leq (d-\epsilon)|B|\}$ be the "bad set". If $|A| > \epsilon|X|$, then (A,B) is a pair violates the regularity condition of (X,Y). □

Now we state (the modern version of) the regularity lemma: Every sufficiently large graph can be partitioned into not too many equal parts so that most pairs of the parts are ϵ-regular.

Lemma 17.3 (Szemerédi 1978)**.** *For every $\epsilon > 0$ and positive integer m, there exist integers N and M such that, whenever G is a graph on $n > N$ vertices, the vertices of G can be partitioned into $k+1$ parts $\{X_0, X_1, ..., X_k\}$ such that*

- $m \leq k \leq M$.
- $|X_0| \leq \epsilon n$.
- $|X_i| = |X_j|$ *for all* $1 \leq i < j \leq k$.
- (X_i, X_j) *is ϵ-regular for all but $\epsilon\binom{k}{2}$ of the pairs $\{i,j\} \in \binom{[k]}{2}$.*

Note: It is a standard trick to combine ϵ and m into one constant, and N and M into one; but we feel in this case fewer variables does not help to clarify the idea.
Note: The only purpose of X_0 is to make all the other parts of the same size. In an equivalent form of the lemma one may drop X_0 and require $||X_i| - |X_j|| \leq 1$ for all i, j.

Note: In most applications, we start from G, get a partition of the vertices into at most M parts, then look at only edges between regular pairs. As we have seen, the regular pairs nicely resemble random bipartite graphs, and we can often find nice properties inspired by random graphs. That means we are disregarding the edges (i) incident to X_0; (ii) between the irregular pairs; (iii) inside each X_i. The conditions of the lemma ensure that there are not so many such edges. In particular, $k \geq m$ ensures that there are not many edges of kind (iii).
We first give some typical applications of the lemma, then we are going to very briefly sketch a proof of the lemma.
The lemma gives a simple proof to Erdös-Stone theorem. Choose ϵ and m appropriatelty and any big graph G has a partition according to the lemma. When there are more than $(1 - 1/p + \epsilon_0)n^2/2$ edges in the graph, and n big enough, the number of edges between the regular pairs is still enough so that more than $(1 - 1/p)\binom{k}{2}$ of the pairs are regular with positive density, more than some pre-fixed $d > 0$. Now by Turán's theorem, we can find $p+1$ parts among which any pair is regular with density at least d. Now when n is large enough, and if we daydream about randomly picking edges with probability d, the probability that we find a $T(q, p + 1)$ tends to 1. This is completely sketchy. We are going to make it more precise.

Definition 17.3. *$G = (V, E)$ a graph, $\epsilon, d > 0$, $P = \{V_0, V_1, ..., V_k\}$ a partition of V. The* reduced graph *$R_{G,\epsilon,d,P} = R$ is defined to be $R = ([k], E')$, where $ij \in E'$ is (V_i, V_j) is an ϵ-regular pair with density at least d.*

Lemma 17.4. *Let $d > 0$ and $\epsilon > 0$ be two (small) constants, and $d+\epsilon < \alpha < 1$. Then there are N and M such that any graph with $n > N$ vertices and at least $\alpha n^2/2$ edges has a partition $P = \{V_0, V_1, ..., V_k\}$ where $k < M$, $|V_i| < \epsilon n$, $|V_i| = |V_j|$ for $0 < i < j \leq k$, and $R_{G,\epsilon,d,P}$ has at least $(\alpha - d - \epsilon)k^2/2$ edges.*

Proof. Let ϵ_0 T.B.D., and $m = \lceil 1/\epsilon_0 \rceil$. Apply the regularity lemma on ϵ_0 and m and let N_0 and M be the constants given by the lemma. Now G has more than N vertices, let P be a good partition given by the lemma. Clearly $k < M$, $|V_i| < \epsilon_0 n$, $|V_i| = |V_j|$ for $0 < i < j$. If $R_{G,\epsilon_0,d,P}$ has less than $(\alpha - d - \epsilon)k^2/2$ edges, we may conclude that the original graph has at most

- Edges incident to V_0. $\epsilon_0 n \cdot n$.
- Edges inside each V_i, $i > 0$.

$$k\left(\frac{n}{k}\right)^2 = \frac{n^2}{k} \leq \frac{n^2}{m} \leq \epsilon_0 n^2.$$

- Edges between the irregular pairs.

$$\epsilon_0 \binom{k}{2} \left(\frac{n}{k}\right)^2 \leq \frac{\epsilon_0 n^2}{2}.$$

- Edges between the regular pairs with density less than d.

$$\binom{k}{2} d \left(\frac{n}{k}\right)^2 \leq \frac{dn^2}{2}.$$

- Edges between the regular pairs with density at least d.

$$(\alpha - d - \epsilon)\frac{k^2}{2}\left(\frac{n}{k}\right)^2 \leq (\alpha - d - \epsilon)\frac{n^2}{2}.$$

Now choose $\epsilon_0 = \epsilon/6$, a contradiction. Finally notice that $R_{G,\epsilon_0,d,P}$ is a subgraph of $R_{G,\epsilon,d,P}$.

□

Consider a fixed graph H and any small probability $d > 0$. A random graph where each edge is picked with probability d almost surely contains a copy of H, when the number of vertices $n \to \infty$. The following lemma says that the regular pairs achieve the same effect.

Definition 17.4. *Let G be a graph, t a positive integer, define $G(t) := (V', E')$, where*

$$V' = V \times [t],$$

and

$$E' = \{\{(u,i),(v,j)\} : 1 \leq i,j \leq t, uv \in E\}.$$

i.e., we change each vertex into t copies of independent vertices, and each edge to t^2 copies. e.g., the Turán graph $T(tp,p)$ is $K_p(t)$.

Lemma 17.5. *Given $d > \epsilon > 0$, a graph R on $[k]$, positive integer t. Let H be a subgraph of $R(t)$ with maximum degree Δ. Let G be a graph with vertex partition $V_1, ..., V_k$ such that whenever ij is an edge in R, (V_i, V_j) is an ϵ-regular pair with density at least d. If*

$$\epsilon < \frac{(d-\epsilon)^\Delta}{\Delta + 2} \text{ and } \frac{(d-\epsilon)^\Delta}{\Delta+2}|V_i| > t - 1 \text{ for all } i,$$

then G contains H as a subgraph.

Note that for any $d > 0$ and Δ, we can pick ϵ small enough (say $\epsilon < \min((d/2)^\Delta/(\Delta+2), d/2)$) to satisfy the first condition.

Proof. (Very sketchy) Let the vertices of H be u_1, u_2, ..., u_{n_0}, we are going to find them out in G one by one.
Each u_i belongs to a part that was generated by a vertex x_i of R. We keep a set of candidates C_i for u_i. In the beginning, for each i, $C_i = V_{x_i}$, the part corresponding to x_i in G.
In the i-th step, we will fix u_i. Consider its neighbours u_j for $j > i$, look at their candidate sets C_j. Call a candidate $v \in C_i$ (which is a vertex of G) bad if v has less than $(d-\epsilon)|C_j|$ neighbours in C_j for some j. Use Fact 17.2, there are at most $\Delta(d-\epsilon)|C_i|$ bad vertices, so we find a good vertex, and then shrink each C_j to its neighbours of v, which is still more than $(d-\epsilon)$ proportion of the size before shrinking.
It is left to check: the conditions of this lemma ensures that all the pairs we consider has at least ϵ proportion of the respective parts, so Fact 17.2 is applicable. □

Now we make the proof of Erdös-Stone theorem look better.

Proof. (of Theorem 16.9) Let G be a graph with n vertices and $(1-1/p+\epsilon_0)n$ vertices. Let $\alpha = 1-1/p+\epsilon_0$ and we will find small ϵ and d that will only depend on the constants ϵ_0 and $T(q, p+1)$. First let $d < \epsilon_0/3$ and $\epsilon < \epsilon_0/3$, so by Lemma 17.4, when n is big enough, there is a good partition P with $k < M$ parts such that $R_{G,\epsilon,d,P}$ has more than $(1-1/p+\epsilon_0-d-\epsilon)k^2/2 > (1-1/p)k^2/2$ edges. By Turán's theorem, it has a $p+1$ clique. i.e., $p+1$ parts that are pairwise ϵ-regular with density at least d.
Now we turn to Lemma 17.5. Make ϵ small enough (but still a constant) to satisfy the lemma. $k \leftarrow p+1$, $R \leftarrow K_k$, $t \leftarrow \lceil q/(p+1) \rceil$, $H \leftarrow T(q, p+1)$. Now G contains $T(q, p+1)$ as soon as $|V_i|$ is bigger than a certain constant. Note that there are at most M (which is also a constant) parts to begin with. So we are done as soon as n is more than a constant. □

Recall that the Ramsey number $r(n_0, n_0)$ grows exponentially in terms of n_0. For any graph H, we may ask the graph Ramsey number: What is the largest n such that when edges of K_n is 2-coloured, there is no monochromatic H? Here is one of the most celebrated results in graph Ramsey theory, also one of the first deep applications of the regularity lemma: When the degree of H is bounded, the Ramsey number is linear in n_0.

Theorem 17.6 (Chvátal - Rödl - Szemerédi - Trotter 1983)**.** *For any integer $\Delta > 0$, there is a constant $c > 0$ such that, whenever H is a graph on n_0 vertices and $\Delta(H) \leq \Delta$ and G is a graph on $n > cn_0$ vertices, then G or $\overline{G}$ contains H as a subgraph.*

Proof. (Sketch) Let ϵ T.B.D. (the important thing is that they will only depend on Δ) and $m = \lfloor 1/\epsilon \rfloor$ and apply regularity lemma, there are N and M given by the lemma. We pick $c > \max(M, N)$ T.B.D. (also will only depend on Δ).
Let G be a graph on cn_0 ($> N$) vertices, there is a good partition by the regularity lemma. There are at least $(1-\epsilon)\binom{k}{2}$ regular pairs. When ϵ is small enough, say $\epsilon = 2^{-4\Delta}$, the number of regular pairs is more than $(1 - 1/2^{2\Delta})k^2/2$. By Turán's theorem, there are $q > 2^{2\Delta}$ parts who are pairwise regular.
It is clear (exercise) that if (X, Y) is ϵ-regular with density d in G, then it is also ϵ-regular with density $1 - d$ in $\overline{G}$. Among the q pairs, colour a pair yellow if it has density at least 0.5 in G, otherwise it is a regular pair with density at least 0.5 in $\overline{G}$ and we colour it blue. By our bound on the Ramsey number $r(\Delta, \Delta)$, w.l.o.g., there are $\Delta + 1$ pairs X_1, X_2, ..., $X_{\Delta+1}$ that are pairwise regular with density at least 0.5.
Now use $d \leftarrow 0.5$ in Lemma 17.5 and make ϵ even smaller than the one from the lemma that only depends on 0.5 and Δ. $k \leftarrow \Delta + 1$, $R \leftarrow K_k$, $t \leftarrow n_0$. Now let $c > (M+1)(\Delta+2)/(d-\epsilon)^{\Delta}$, which only depends on Δ. It is easy to check that the lemma implies H is a subgraph of G. □

We only outline a proof of the regularity lemma. Start from any equipartition into m parts. For any partition $\mathcal{P}$ of the vertices into the exceptional V_0 and other k equal parts, define

$$f(\mathcal{P}) = \frac{1}{k^2} \sum_{1 \leq i < j \leq k} d^2(V_i, V_j).$$

Since each density is at most 1, clearly $f(\mathcal{P}) \leq 1/2$ for any partition.
The key step is: If there are more than ϵk^2 irredular pairs, then we can refine $\mathcal{P}$ to another partition $\mathcal{Q}$ such that $\mathcal{Q}$ has about $k4^k$ parts, and

$$f(\mathcal{Q}) \geq f(\mathcal{P}) + \frac{\epsilon^5}{20}.$$

Since $1/2$ is an upper bound, this process cannot take more than $10/\epsilon^5$ steps.

We note that the bound on the number of parts in the regularity lemma thus gotten is more than $4^{4^{4^{\cdot^{\cdot^{\cdot^4}}}}}$, where the tower is of height $10/\epsilon^5$.

> I understand internet, that's just a graph, I can model it. But the computer, programming languages, how to search for information, I don't know.
>
> – Endre Szemerédi